

DIFFERENTIAL EXPRESSION OF EIGHT TRANSCRIPTS AND THEIR ROLES IN THE  
CUTICLE OF THE BLUE CRAB, *CALLINECTES SAPIDUS*

Lindsay Faircloth

A Thesis Submitted to the  
University of North Carolina Wilmington in Partial Fulfillment  
of the Requirements for the Degree of  
Master of Science

Department of Biology and Marine Biology

University of North Carolina Wilmington

2006

Approved by:

Advisory Committee

---

---

---

Chair

Accepted by

---

Dean, Graduate School

This thesis has been prepared in the style and format  
consistent with the journal  
Comparative Biochemistry and Physiology, Part B

## TABLE OF CONTENTS

ABSTRACT.....	iv
ACKNOWLEDGEMENTS.....	vi
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
INTRODUCTION.....	1
METHODS.....	7
Sequence Determination of Cloned Transcripts.....	7
Tissue Collection, RNA Isolation, and cDNA Synthesis.....	8
DIG-labeled Probe Synthesis and Northern Blots.....	10
Quantitative PCR.....	11
<i>In situ</i> Hybridization.....	12
Nomenclature.....	13
RESULTS.....	13
Arthrodial Membrane Proteins.....	13
Calcifying Cuticle Proteins.....	18
Expression Data.....	22
<i>In situ</i> Hybridization.....	25
DISCUSSION.....	30
LITERATURE CITED.....	38
APPENDIX.....	41

## ABSTRACT

Biom mineralization in the exoskeleton of decapods is under the control of proteins deposited in each of the cuticle layers by the hypodermis. The epicuticle and exocuticle are deposited before ecdysis, but their mineralization is delayed until after the molt. The endocuticle, however, mineralizes as it is deposited postecdysis. Furthermore, cuticle at the joints known as arthrodial membrane, consists of both pre- and post-ecdysial layers, but does not mineralize at all. Therefore, obtaining sequences and conducting expression analyses for many cuticular proteins may provide insights into the process of mineral formation. This study examines the expression patterns of eight new cuticle protein genes from *Callinectes sapidus* across the molt cycle in both calcifying and non-calcifying cuticle hypodermis using quantitative PCR, Northern blots, and *in situ* hybridization. Four of the transcripts were found only in calcifying hypodermis and are designated CsCP followed by the molecular weight of the predicted mature protein. CsCP6.1 is expressed post-molt and contains a Rebers-Riddiford (RR) motif, suggesting that it binds to chitin in the endocuticle. CsCP14.1 also contains an RR motif. Interestingly, it is the RR-1 variant common in non-calcifying cuticle proteins. The other two transcripts from calcifying hypodermis do not code for RR proteins, but both contain three copies of a different cuticle motif. One of these, CsCP19.0, is expressed only post-molt while the other, CsCP15.0, is present both before and after ecdysis. The other four transcripts characterized in this study are found only in arthrodial membrane hypodermis and are designated CsAMP. They all contain RR-1 motifs and so presumably bind chitin in the non-calcifying cuticle. CsAMP9.3 is most likely an exocuticle constituent since it is expressed only during pre-molt. The other three transcripts are present both before and after ecdysis. Though the specific timing of gene expression differs, all these proteins may be necessary structural

elements common to all non-calcifying arthrodial cuticle layers. One of these, CsAMP16.3, contains the RGD cell-attachment motif. It could be involved in anchoring the hypodermis of pore canals to chitin fibers.

## ACKNOWLEDGEMENTS

My thanks go to Drs. Guy Caldwell and Julie Olson whose instruction and enthusiasm sparked my continued interest in molecular biology. Their constant support and encouragement have made me extremely grateful.

I would also like to thank fellow graduate students, Shannon Modla, Anna Wynn, and Al Nyack, as well as Dr. Francie Coblenz for their expertise, assistance, and good humor that got me through this research, sanity intact. This thesis would also not have been possible without the assistance of Mark Gay who was never too busy to answer a question.

Special thanks go to my parents and my sister whose support I can always count on, and to Joe Rizzardi who listened to hours of graduate student strife with absolutely no complaints. Without their love and support, I could not have accomplished nearly as much as I have.

Finally, I would like to thank my committee for their guidance, criticism, and encouragement. They never doubted my ability to do better and have helped make this work the best that it can be.

## LIST OF TABLES

	Table	Page
1.	Primer sequences and annealing temperatures used for synthesis of DIG probes and qPCR .....	9
2.	Summary of expression and sequence data with proposed functions for all eight transcripts .....	31

## LIST OF FIGURES

Figure	Page
1. Eleven of the known RR-1-containing proteins from crustaceans as well as five from this study are shown in this alignment: CpCP/AM11.14 (P81575), CpAM11.59 (P81576), HaAMP1A (P81384), DD9A (BAA90875), DD9B (BAA90876), CpAM11.99 (P81577), CpAM12.39 (P81578), CpAM12.74 (P81579), HaAMP5 (P81389), CsAMP8.1 (AAV28476), and CsAMP6.0 (AAV28477). HaAMP1A is representing the group that includes HaAMP1B (P81385), HaAMP2 (P81386), HaAMP3 (P81387), and HaAMP4 (P81388). The five proteins from this study are underlined. The RR-1 and upstream consensus sequences are shown below the alignment. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids .....	16
2. Alignment of virtually translated CsCP6.1 with its most homologous sequences in both Swiss-Prot and NCBI databases: <i>C. pagurus</i> CpCP5.75 (P81589) and <i>P. clarkii</i> CAP-2 (BAD16776). The RR-1 consensus is shown below the alignment in bold. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids .....	20
3. Alignment of virtually translated CsCP15.0 and CsAMP19.0 (a shown) with their most homologous sequences in both Swiss-Prot and NCBI databases: <i>C. pagurus</i> CpCP14.99 (P81583), CpCP18.76 (P81584) and HaCP20.2 (a shown) (Q7M496). The post-molt 18 amino acid motif is shown and the repeated regions are denoted by dashes below the alignment. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids .....	21
4. qPCR graphs and Northern blots for (a) CsAMP16.5, (b) CsAMP16.3, (c) CsAMP13.4, and (d) CsAMP9.3. A vertical bar separates pre-molt stages from post-molt time points. Data for individual crabs is indicated by the open or closed circles and the bars indicate the range, or variation, in the data. (Note: The scales are not the same on each graph and only one sample from the arthrodial e-D <sub>2</sub> stage was measured.).....	23



5. qPCR graphs and Northern blots for (a) CsCP6.1, (b) CsCP19.0, (c) CsCP15.0, and (d) CsCP14.1. A vertical bar separates pre-molt stages from post-molt time points. Data for individual crabs is indicated by the open or closed circles and the bars indicate the range, or variation, in the data. (Note: The scales are not the same on each graph and only one sample from the arthrodial e-D<sub>2</sub> stage was measured.).....26
  
6. Sections of the walking leg joint from pre-molt stage D<sub>2</sub> (a-d), 48 h post-molt (e), and 12 h post-molt (f) crabs. Only arthrodial hypodermis is shown in (f). Sections were (a) stained with hematoxylin and eosin, (b) used as a control in *in situ* hybridization in which no probe was hybridized, (c) hybridized with the sense probe for CsCP14.1, (d) hybridized with the anti-sense probe for CsCP14.1, (e) CsAMP16.5, and (f) CsAMP16.3. In sections a-e, hypodermis to the right of the arrowhead synthesizes calcifying cuticle and to the left synthesizes arthrodial membrane cuticle (non-calcifying). Both hypodermis (h) and cuticle (c) are labeled on each section .....28

## INTRODUCTION

Biom mineralization is important in the formation of the hard exoskeletons of crustaceans and occurs within the organic matrix of the cuticle. This organic matrix contains approximately 60% protein and 40% chitin (Horst, 1990). The chitin fibers of the organic matrix can interact with these proteins. Some of the proteins are responsible for the regulation of initiation and termination of mineral formation (Addadi and Weiner, 1985). Specifically, acidic proteins are thought to attract  $\text{Ca}^{2+}$  ions and serve as nucleation centers for  $\text{CaCO}_3$  in the form of calcite in crustaceans. Insights into biom mineralization can be gained by studying when and where these proteins are deposited in the cuticle and their specific characteristics.

The blue crab, *Callinectes sapidus*, calcifies its cuticle through the process of biom mineralization on a regular basis. The cuticle is composed of four layers whose synthesis and deposition is controlled by the underlying layer of epithelial cells (hypodermis). During its lifetime the blue crab goes through many molts in which the cuticle is shed and a new one is synthesized and deposited. This molt cycle has been characterized and separated into pre-molt ( $D_0$ ,  $D_1$ ,  $D_2$ ,  $D_3$ ,  $D_4$ ), ecdysis (E), post-molt ( $A_1$ ,  $A_2$ ,  $B_1$ ,  $B_2$ ,  $C_1$ ,  $C_2$ ,  $C_3$ ), and intermolt ( $C_4$ ) stages (Drach and Tchernigovtzeff, 1967). Each stage is accompanied by specific morphological changes in these cuticle layers.

The initiation of pre-molt ( $D_0$ ) is marked by apolysis, the separation of the hypodermis from the old cuticle (Roer and Dillaman, 1993). There are two pre-exuvial layers that are synthesized and deposited prior to ecdysis. The epicuticle is the outermost layer composed of lipids, proteins, and chitin and is deposited during early pre-molt ( $D_1$ ). During late pre-molt ( $D_2$ ) the chitin-rich exocuticle is deposited. The epi- and exocuticle layers are fully formed during pre-molt, yet are not capable of nucleating  $\text{CaCO}_3$  until 1-3 hours after ecdysis when a post-

ecdysial cuticle alteration (PECA) occurs in the cuticle (Shafer et al., 1995). It has been hypothesized that the nucleating proteins are present during pre-molt, but inactive because they are shielded by large macromolecules that are removed or altered at ecdysis to allow for calcification (Coblentz et al., 1998). The thickest cuticle layer, the endocuticle, begins its deposition as early as 6 hours post-molt and is simultaneously calcified (Dillaman et al., 2005). The membranous layer is deposited last (C<sub>3</sub>) and never calcifies (Roer and Dillaman, 1984). Intermolt (C<sub>4</sub>) is marked by the end of membranous layer deposition and the cessation of calcification.

There are some cuticle types that do not follow this pattern of deposition and calcification. The cuticle at the joints of the appendages (arthrodial membrane), overlying the gills and the inner branchial chamber, and lining the fore and hindgut remain uncalcified to allow for movement, nutrient absorption, and oxygen transfer (Neville, 1975). The arthrodial membrane has the same pattern of deposition as calcified cuticle and this allows for direct comparisons between calcified and non-calcified cuticle (Williams et al., 2003). Just as calcification is not uniform throughout the cuticle, neither is the timing of cuticle deposition (Williams et al., 2003). The inner branchial chamber and gill cuticle layers are all deposited prior to ecdysis (Elliott and Dillaman, 1999).

By looking at gene expression in the hypodermis at different molt cycle stages and in both calcifying and non-calcifying cuticle, the proteins being put into each layer and tissue type can be determined and the differences compared.

Many proteins from crustacean cuticles have been isolated and characterized. Fifteen proteins have been extracted from the calcified exoskeleton of the lobster, *Homarus americanus* (Kragh et al., 1997; Nousiainen et al., 1998) and six from the arthrodial membrane (Andersen,

1998a). Two of the calcified cuticle proteins contain one copy of an 18-residue motif (VxDTPEVAAAKAAFxAAY, where x is any amino acid) that has been reported in post-ecdysial insect cuticle (Andersen, 2000) and will be referred to as the insect-post18 motif. Nine other calcified cuticle proteins contain either two or four copies of another hydrophobic 18-residue sequence (x[L/V][I/V]GPSGIV[T/S]x[D/N]GxN[I/V]Q[V/L], where x represents any residue) that is unique to crustaceans (Kragh et al., 1997; Nousiainen et al., 1998) and will be referred to as the crust18 motif. The frequent occurrence of residues in this motif that contain carboxylate and amide groups indicate that it could have sufficient  $\text{Ca}^{2+}$  affinity to be a nucleating center (Nousiainen et al., 1998). The six arthrodial membrane proteins are not similar to those extracted from the calcified exoskeleton but are similar to each other (Andersen, 1998a). These arthrodial membrane proteins all contain a Rebers-Riddiford (RR) consensus sequence (Rebers and Willis, 2001; Togawa et al., 2004).

The RR consensus sequence is highly conserved among both insect and crustacean cuticle proteins (Rebers and Riddiford, 1988; Andersen, 1998a, 1998b) and has been experimentally shown to bind chitin (Rebers and Willis, 2001; Togawa et al., 2004). There are three variants, RR-1, RR-2, and RR-3 (Andersen, 1998a, 1998b, 2000). The RR consensus is  $\text{Gx}_8\text{Gx}_6\text{YxAxExGYx}_7\text{Px}_2\text{P}$ , where x indicates any amino acid (Rebers and Riddiford, 1988). This sequence along with a conserved upstream region is considered to be the RR-1 variant. RR-1 proteins are generally found in soft, flexible, hydrated cuticles like the arthrodial membranes of crustaceans (Andersen, 1998a). RR-2 proteins contain the residues GFNAVV instead of the two terminal prolines in the RR consensus, and have a more highly conserved upstream region that is different from the same region in RR-1 (Andersen, 1998b). RR-3 is slightly different in the RR consensus and very different from the other two variants in the N-terminal region (Andersen,

2000). RR-2 has not yet been found in crustacean cuticle proteins. RR-3 has been found in a protein from the calcified cuticle of *H. americanus* (Andersen, 1998a) and there is one other *H. americanus* calcified cuticle protein that has a region very similar to RR-3. All the *H. americanus* arthroal membrane proteins contain RR-1 (Andersen, 1998a).

Sixteen cuticle proteins from the crab, *Cancer pagurus*, have also been identified (Andersen, 1999). Eleven were extracted from calcified cuticle, four from arthroal membrane cuticle, and one was found in both. Eight of the calcified cuticle proteins contain copies of the crust18 motif, while two contain copies of the insect-post18 motif. One of the other calcified cuticle proteins contains a truncated RR-1 and the one found in both tissue types contains a complete RR-1 (Andersen, 1999). The proteins from arthroal membrane cuticle all contain the RR-1 consensus and are very similar to each other and to *H. americanus* arthroal membrane proteins. A model has been proposed in which the majority of arthroal membrane proteins are involved in chitin binding via this RR consensus. The RR-containing proteins in calcified cuticle also bind chitin, but hydrophobic proteins that do not contain the RR consensus dominate the space between chitin filaments as opposed to water dominating this space in arthroal membrane cuticles (Andersen, 1999).

Two very similar cuticle protein transcripts, DD9A and B, have been isolated from the shrimp, *Marsupenaeus japonicus* (Watanabe et al., 2000). These two transcripts were expressed post-molt in non-calcifying lateral areas of the tail fan and contain RR-1. It has been hypothesized that these proteins could be involved in preventing calcification of the endocuticle (Watanabe et al., 2000). Two other *M. japonicus* transcripts found in the calcifying medial region of the tail fan were characterized: crustocalcin (formerly DD4) and DD5 (Endo et al., 2000, 2004; Ikeya et al., 2001). Crustocalcin was expressed as early as 2 hours post-molt and

has a RR region as well as a  $\text{Ca}^{2+}$  binding region shown to nucleate  $\text{CaCO}_3$  *in vitro* (Endo et al., 2004). There is very little sequence upstream from the RR region so this protein cannot be considered either RR-1, RR-2, or RR-3. Crustocalcin was detected immunohistochemically in the endocuticle, but not in the exocuticle, indicating that different molecular mechanisms may control calcification in these layers. DD5 was expressed in post-molt hypodermis and contains tandem repeats that include the RR consensus (Ikeya et al., 2001), but not any of the conserved upstream regions. This tandemly repeated RR region may be involved in cross linking chitin fibers in the cuticle, and the uniform size of the repeats may reflect the regular spacing of chitin fibrils (Ikeya et al., 2001).

Two calcified cuticle proteins, CAP-1 and CAP-2, from the crayfish, *Procambarus clarkii*, have also been characterized (Inoue et al., 2001, 2003, 2004). Both transcripts were expressed post-molt and the translated proteins contain a RR consensus. The N- and C-terminal regions are very short and cannot be considered RR-1, RR-2, or RR-3. These proteins are thought to be involved in the nucleation of  $\text{CaCO}_3$  in the endocuticle (Inoue et al., 2001). Both were shown to bind chitin and  $\text{Ca}^{2+}$  *in vitro*, but CAP-1 has two  $\text{Ca}^{2+}$ -binding regions while CAP-2 only has one (Inoue et al., 2004). These proteins are 44% identical with the major differences being in the C-terminal region.

There have been 15 transcripts described in *C. sapidus*, thirteen from calcifying cuticle hypodermis and two from arthrodial membrane hypodermis. Two of the calcifying cuticle transcripts are very similar to the CAP proteins, but are expressed during pre-molt (Wynn and Shafer, 2005). These proteins are hypothesized (Wynn and Shafer, 2005) to act as nucleating agents that are incorporated into the pre-exuvial layers and remain inactive until after the post-ecdysial cuticle alteration (PECA) (Shafer et al., 1995). They, like the CAP proteins, contain

short, highly acidic N- and C-terminal regions along with the RR consensus. The other eleven calcifying cuticle transcripts were found to contain copies of the crust18 motif (Kennedy, 2004). The two arthroal membrane transcripts were found to contain the RR-1 consensus (Wynn and Shafer, 2005). They are expressed both pre- and post-ecdysis suggesting that they are structurally important components of all arthroal membrane cuticles.

All of the crustacean cuticle proteins, except those from *C. sapidus*, were put into an alignment-based dendrogram to illustrate degrees of similarity among the proteins of different species (Shafer and Faircloth, 2006). Several distinct clusters emerged. One contains all the non-calcifying cuticle RR-1 proteins. A second consists of four RR-containing proteins from calcified cuticle that are thought to be involved in mineral nucleation (CAP-1, CAP-2, crustocalcin, and CpCP5.75). Another consists of proteins from calcified cuticle that have the crust18 motif. There is one cluster of calcified cuticle proteins that all contain three copies of the insect-post18 motif. The last cluster consists of proteins from calcifying cuticles whose functions are as yet unclear (Shafer and Faircloth, 2006). These clusters allow a directed search for the homologs of these proteins in *C. sapidus* to be undertaken. This search was conducted by mining a *C. sapidus* expressed sequence tag (EST) database (Coblentz et al., 2006) for cuticle proteins. Eight cuticle proteins were found and have been put into their respective clusters based on the alignment (Shafer and Faircloth, 2006). In this study, their patterns of expression were determined and compared to cuticle proteins from other crustaceans and some insects. These expression data coupled with sequence homology analyses will lead to a better understanding of the control of mineralization and the differences between calcified cuticle and non-calcified arthroal membrane.

## METHODS

### Sequence Determination of Cloned Transcripts

A recently completed expressed sequence tag (EST) database includes sequences from both pre- and post-molt hypodermis (Coblentz et al., 2006). cDNA clones selected from this database were grown overnight in Terrific Broth (TB) (MO BIO) with 10% glycerol and carbenicillin (100 µg/mL). The plasmids were purified using the Wizard Plus SV Miniprep kit (Promega). The insert was amplified by PCR with primers specific to the T7 and SP6 promoter regions on the plasmid (Table 1). All PCR reactions consisted of 2.5 µL 10X BD Advantage 2 Buffer (BD Biosciences), 0.5 µL dNTP mix (10 mM) (Invitrogen), 1 µL of each primer (10 mM), 0.25 µL 50X BD Advantage 2 Polymerase mix, 1 µL DNA, and 18.75 µL nuclease-free water. The PCR reactions began with a hot start at 94°C for 1 min followed by a total of 30 cycles at 94°C for 30 s, at the primer-specific annealing temperature (Table 1) for 1 min, and 72°C for one min.

To verify amplification of the insert, 5 µL of each PCR product was assessed by electrophoresis on a 1.6% agarose gel with 6 µg ethidium bromide at 70 V. The remaining product was purified for sequencing using the Wizard SV Gel and PCR Clean Up System (Promega). Sequencing reactions contained 1 µL of DNA, 2 µL Big Dye Terminator Ready Reaction Mix v 3 (Applied Biosystems), 2 µL dilution buffer, 3.4 µL nuclease-free water, and 1.6 µL of either T7 or SP6 vector primers (1 µM). Reaction conditions were 94°C for 45 s, 50°C for 45 s, and 60°C for 4 min for 25 cycles. The sequencing reaction products were purified by centrifugation through G-50 Sephadex at 2,000 x g for 2 min and analyzed on an ABI 3100 DNA sequencer (Applied Biosystems). Forward and reverse sequences were examined and vector sequences trimmed using Chromas software v 2.13 (Technelysium).



The open reading frames (ORFs) were virtually translated using VectorNTI software v 10.0 (Invitrogen) and signal sequences were predicted (ExPASy-SignalP 3.0). Signal sequences were trimmed before the molecular weights and isoelectric points were predicted (ExPASy-ProtParam Tool). The mature protein sequences were then compared to the Swiss-Prot and NCBI databases and the results were aligned in VectorNTI (v 10.0), which utilizes the ClustalW algorithm.

#### Tissue Collection, RNA Isolation and cDNA Synthesis

Adult *C. sapidus* organisms were obtained from a “soft shell shedding” operation in Kill Devil Hills, NC, in May 2005. Hypodermal tissues were collected for RNA isolation from pre-molt stages that were determined according to the method of Drach and Tchernigovtzeff (1967). Post-molt tissues were obtained by observing ecdysis and sacrificing the crabs at designated times thereafter. Arthrodial membrane hypodermis was dissected from the joint proximal to the carpus segment of both chelipeds. Mid-dorsal hypodermis was dissected from the area above the cardiac chamber so as to be free from muscle attachment sites and other tissue. Hypodermis tissue was stored in RNA Later (Ambion) at -20°C.

Total RNA was extracted from these tissues using the spin-column RNeasy Protect Mini Kit (Qiagen), with the following modifications to increase yield and quality. Tissue was homogenized in 1 mL TRIzol (Invitrogen). RNA was eluted from the column with 30 µL nuclease-free water (Ambion), and the eluate was passed through the column again to increase the yield. Accurate measures of RNA quality and quantity were determined using the Agilent 2100 BioAnalyzer and RNA 6000 Nano Assay Kit.

Table 1

Primer sequences and annealing temperatures used for synthesis of DIG probes and qPCR

Primer Name	Primer Sequence 5' - 3'	Location	Temp <sup>a</sup> (°C)
T7	TAATACGACTCACTATAGGG	vector	55
SP6	ATTTAGGTGACACTATAAGAA	vector	55
CP15.0ProbeF	CTTGCTCGGCTTTACCAGTC	69-89	55
CP15.0ProbeR	GCCTCTGGTAGGTGTCGAAG	497-517	55
AMP13.4ProbeF	GGGTCTGTGAGCAGGAGAAG	498-518	52
AMP13.4ProbeR	CATTGGATTGTGTCCAGTGC	915-935	52
CP6.1ProbeF	CACCAACATGAAGGTCCTGA	17-37	54
CP6.1ProbeR	TAAACATGAGGTCGGTGTCTG	341-361	54
AMP9.3ProbeF	AGAACATCAACGAGGGCAAC	105-125	54
AMP9.3ProbeR	AGGGACAGCCATCCTCCTAT	475-495	54
CP14.1ProbeF	GCACTCAGCATCATGCACTT	6-26	54
CP14.1ProbeR	CAAAGGCGATCTGGTCTAGC	381-401	54
CP19.0ProbeF	TCAGACCCCTAAGTGGATGG	254-274	54
CP19.0ProbeR	GTGCGGCTTAGAAGAACCTG	657-677	54
AMP16.5ProbeF	CCTCACGTAGCCATCCTGTT	171-191	54
AMP16.5ProbeR	AGGCTGTTCGGAGCTCATTA	610-630	54
AMP16.3ProbeF	CGGAGACGGCAACTACAAC	195-215	55
AMP16.3ProbeR	CTATGGACGCCGTGTATGGT	650-670	55
CP15.0QPCR	ACTCCAGCAACACACTAACATCATG	13-38	60
CP15.0QPCR	CAGCACTGCCAAGACAACCA	64-84	60
AMP13.4QPCR	GGGACTGCTCGACACCTGTAA	569-590	60
AMP13.4QPCR	GTCGCCTTATGACGTGGATGT	620-641	60
CP6.1QPCR	AGGACGCCTTCGGCAACT	161-179	60
CP6.1QPCR	CACCTGCGTCAGCGATGTACT	211-232	60
AMP9.3QPCR	GCAACGGCGTCTTCCAGTAC	15-35	60
AMP9.3QPCR	CCACTACAATGCCGTTGTCTGT	65-87	60
CP14.1QPCR	TTCGTGGCCAACGAGAATG	168-187	60
CP14.1QPCR	CCACAGGGAGCAGAGGAGACT	219-240	60
CP19.0QPCR	GCCCTCTTGCTTCCAACGT	21-40	60
CP19.0QPCR	GGCGTTCAGGAAATGATTCTTG	121-143	60
AMP16.5QPCR	GGCGAGCCACAGTTCTCTTAAT	310-332	60
AMP16.5QPCR	CAGGTACAGCAGGTATGAGAGAACA	363-388	60
AMP16.3QPCR	GGCCAGACCAACCAAGCA	74-92	60
AMP16.3QPCR	CGAGGGTGCCATCAGGAA	128-146	60

<sup>a</sup> Annealing temperatures used in PCR or qPCR.

RNA as above, and nuclease-free water to a volume of 24  $\mu$ L. This reaction mixture was heated to 70°C for 5 min and held at 48°C. A mixture of 8  $\mu$ L 5X First-Strand Buffer (Invitrogen), 4  $\mu$ L DTT (0.1 M), and 2  $\mu$ L RNase OUT Recombinant Ribonuclease Inhibitor (40 U/ $\mu$ L) (Invitrogen) was warmed to 48°C before being added to each reaction. After 2 min at 48°C, 1  $\mu$ L (200 units) of Superscript II reverse transcriptase (Invitrogen) was added to each reaction. The thermocycler parameters were 48°C for 50 min, 70°C for 15 min, and 37°C for 25 min. Five minutes into the 37°C incubation, 1  $\mu$ L (2 units) of *E. coli* RNase H (Invitrogen) was added.

#### DIG-labeled Probe Synthesis and Northern Blots

Primer sets amplifying a PCR product between 400 and 500 bp in length were chosen for each transcript using Primer3 software ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). Probe sequences were checked for specificity by comparing to both the *C. sapidus* EST database and NCBI's database using BLASTn. They were also scanned for the absence of SpeI restriction sites using VectorNTI v 10.0. PCR products were obtained for each primer set using appropriate cDNA. Annealing temperatures used for each primer set are shown in Table 1.

Using the pGEM-T Easy Vector System (Promega) each PCR product was ligated into the pGEM-T Easy vector and cloned into high efficiency competent JM109 *E. coli* cells. Transformed cells were plated onto LB agar plates containing carbenicillin (100  $\mu$ g/mL) by using sterile beads to evenly distribute 100  $\mu$ L of the cell culture. Plates were incubated at 37°C for 24 h. PCR was performed on individual colonies using SP6 and T7 primers to verify the presence of the inserts. The PCR products were then sequenced to determine the direction of the insert. Colonies containing sense (non-complementary) and anti-sense (complementary) inserts were grown in TB broth with 10% glycerol and carbenicillin (100  $\mu$ g/mL) at 37°C overnight.

The plasmids were purified using the Wizard Plus SV Miniprep kit (Promega). One microgram of purified plasmid was linearized with SpeI restriction enzyme (Roche) following the manufacturer's protocol and purified using the Wizard SV Gel and PCR Clean Up System (Promega). Riboprobes were transcribed from the T7 promoter and labeled with digoxigenin (DIG) using the DIG RNA labeling kit (Roche).

Labeled probes were used in the Northern blot analyses. One microgram of each RNA sample was denatured in RNA loading dye (GenHunter) and fractionated on a 1.0% agarose gel with 2% formaldehyde. The RNA migrated at 35 V for 3-4 h before the gel was examined under UV light. The RNA was transferred by capillary action onto an Immobilon-NY+ membrane (Millipore) with 20X SSC (3 M sodium chloride, 0.3 M sodium citrate, pH 7.0) according to the manufacturer's instructions. The RNA was UV cross-linked to the membrane and hybridized overnight at 68°C with DIG-labeled probes diluted 1:10 with DEPC-treated water. Probe binding was detected using an anti-DIG-alkaline phosphatase (AP) conjugate and CSPD, an alkaline phosphatase-activated substrate. Chemiluminescence was detected by 5-10 min exposure to Biomax XAR X-ray film (Kodak).

#### Quantitative PCR

Quantitative PCR (qPCR) was performed using a real-time PCR system (Applied Biosystems, model 7500). Three independent RNA samples were collected for each tissue-time point combination. cDNA was then made from a known amount of RNA (4µg). This allowed all samples to be normalized to comparable RNA concentrations. qPCR primer sets with amplicons between 70 and 130 base pairs were designed for each transcript using Primer Express software (Applied Biosystems) (Table 1). qPCR reactions contained 12.5 µL SYBR Green PCR Master Mix (Applied Biosystems), 1 µL cDNA, 7.5 µL water, and 2 µL of each primer (200

nM). One cDNA sample made from 5 h mid-dorsal RNA was diluted (1.0, 0.1, 0.01, and 0.001) and quantified using CP19.0QPCR and R primers to check for linearity between the log concentration and critical threshold cycle number. All other sample values reported are relative to this one. Data were analyzed using the 7500 System Sequence Detection Software (Applied Biosystems). Thermal dissociation curves for each reaction verified amplification of only one PCR product.

#### In situ Hybridization

Entire joints proximal to the carpus segment of the third walking leg and branchial chamber cuticle were isolated and fixed in a modified Carnoy's fixative (85% EtOH, 10% formaldehyde, 5% glacial acetic acid). Tissues taken from pre-molt time points were decalcified (10% EDTA, 0.1 M Tris HCl, pH 7.4) for two weeks and the exuviae were removed. Pre- and post-molt tissues were embedded in paraffin, cut into 9  $\mu$ m sections, and mounted on Superfrost Plus slides (Fisher). Sections were de-waxed in toluene, rehydrated in a descending acetone series, and post-fixed (4% paraformaldehyde in PBS, pH 7.38) for 20 min. All hybridization solutions and washes were made with DEPC-treated water. Pre-hybridization consisted of several washes at room temperature: 10X TBS (50 mM Tris, pH 7.5, 150 mM NaCl) for 5, 10, and 10 min; 200 mM HCl for 10 min; 10X TBS for 5, 10, and 10 min; acetic anhydride (0.01 M Tris, pH 8, 0.5% acetic anhydride) for 10 min with agitation; 10X TBS as above; and proteinase K (10  $\mu$ g/mL) in TE (10 mM Tris, pH 8, 1 mM EDTA) for 15 min. Finally, three more 10X TBS washes were done at 4°C. The same probes used for Northern blots were used for *in situ* hybridization, but were diluted 1:300 in hybridization buffer (2X SSC, 10% dextran sulfate, yeast tRNA (10 $\mu$ g/mL), 0.02% SDS, 50% formamide). Coverslips were placed onto the slides, held at 95°C for 4 min, sealed with rubber cement, and hybridized overnight at 55°C.

After hybridization, slides were washed once with 2X SSC; three times in 50% formamide with 1X SSC at 55°C for 20 min each; two times in 1X SSC at room temperature for 15 min each; and three times in 1X TBS for 5 min each. They were then treated with blocking agent (Roche) with 10% fetal calf serum and 1% sheep serum for 1 h. An anti-DIG-AP conjugate (Roche) was diluted 1:1000 in the blocking solution and placed on sections for 1 h at room temperature. Sections were washed four times in 1X TBS for 5, 10, 10, and 10 min. They were then equilibrated for 2 min in AP buffer (100 mM Tris, pH 9.5, 50 mM MgCl<sub>2</sub>, 100 mM NaCl). The equilibrated sections were incubated in 0.2 µm-filtered NBT/BCIP substrate solution (Roche) approximately 18 h at 4°C. Slides were mounted with Kaiser's glycerine jelly and the colored product indicating the location of the specific transcript was observed by light microscopy and photographed. Some sections were stained with hematoxylin and eosin in order to discern between arthrodial membrane and cuticle to be calcified in the tissue sections (Williams et al., 2003).

#### Nomenclature

All the transcripts were submitted to GenBank using names generated by designating the species, *C. sapidus* (Cs) followed by a tissue specification of CP if found in hypodermis synthesizing cuticle that will calcify or AMP if found in hypodermis that secretes the arthrodial membrane, followed by the predicted molecular weight in kilodaltons.

## RESULTS

#### Arthrodial Membrane Proteins

Clones were selected from the *C. sapidus* EST database (Coblentz et al., 2006) by searching the annotations for the word “cuticle”. This enabled the discovery of *C. sapidus* proteins similar to known cuticle proteins of other crustaceans (Shafer and Faircloth, 2006).

Eight sequences of clones from the EST library were found to contain complete ORFs. They were virtually translated, the signal peptides were cleaved, and the Swiss-Prot and NCBI databases were searched to confirm and/or update the EST database annotations. Expression studies revealed four transcripts expressed solely in non-calcifying arthroal membrane hypodermis (See Expression Data), and these were named CsAMP16.5, CsAMP16.3, CsAMP13.4, and CsAMP9.3. All of these *C. sapidus* transcripts contain RR-1 and were put into an alignment with 11 of the 15 RR-1 proteins (Figure 1). Five *H. americanus* arthroal membrane proteins that are 90% identical and considered to be isoforms (Andersen, 1998a) are all represented in the alignment by HaAMP1A. All fifteen RR-1 proteins were aligned to determine conservation upstream from the RR sequence, and a 46 amino acid upstream consensus was determined

(DRDAxILTDERxDQGDGNF<sub>x</sub>Y<sub>x</sub>FETSNGIY<sub>x2</sub>KTGTPGSEGQSNMQ).

The first arthroal membrane transcript, CsAMP16.5 (GenBank #DQ288152), is 682 bp with an ORF of 163 amino acids beginning at nucleotide 9 and ending at 497 (Appendix). The polyadenylation signal begins at nucleotide 665. In the virtual translation, the first 15 amino acids are a signal peptide. CsAMP16.5 is most homologous to *C. pagurus* arthroal membrane proteins CpAM11.99 (39% identity) and 12.39 (30% identity) as well as to *M. japonicus* DD9A (32%). When compared to the RR-1 consensus shown at the bottom of the alignment, CsAMP16.5 differs at residue 108 and the final proline is displaced by one residue (Figure 1). CsAMP16.5 is 39% identical to the upstream consensus. This protein has a predicted isoelectric point of 4.94.

The complete sequence for CsAMP16.3 (GenBank #DQ310582) is 700 bp with an ORF of 145 amino acids starting at nucleotide 16 and ending at 496 (Appendix). There is a

polyadenylation signal at position 691. The virtual translation includes a 15 amino acid signal sequence and the RR-1 consensus deviates only twice at residues 94 and 101 where tyrosines are replaced with phenylalanines (Figure 1). BLAST analysis determined that CsAMP16.3 is homologous to the group of similar *H. americanus* arthrodial membrane proteins: HaAMP1A (41% identity), HaAMP1B (40% identity), HaAMP3 and 4 (39% identity) and HaAMP2 (39% identity). The next most significant matches were *M. japonicus* DD9A and DD9B (39% and 36% identity, respectively) and CpAM11.59 and 11.99 (38% and 37% identity, respectively). The N-terminal region is 48% identical to the RR-1 upstream consensus. CsAMP16.3 has a predicted isoelectric point of 6.16.

A third arthrodial membrane protein, CsAMP13.4 (GenBank #DQ288147), is 949 bp with an ORF of 135 amino acids (Appendix). The start codon is at nucleotide 19 and the stop codon at 424. There is a polyadenylation signal at nucleotide 932. The virtually translated protein has a 15 amino acid signal peptide. This protein is homologous to CpAM12.74 (76% identity) and 12.39 (57% identity), previously reported *C. sapidus* arthrodial membrane protein CsAMP8.1 (47% identity), and *M. japonicus* DD9A and DD9B (48% identity for both). CsAMP13.4 shows 46% identity to the RR-1 upstream consensus. The RR-1 consensus itself differs in one amino acid (position 68) and the final proline is displaced by one residue (Figure 1). This protein is acidic with a predicted isoelectric point of 4.61.

CsAMP9.3 (GenBank #DQ288148) is 702 bp with an ORF of 100 amino acids that begins at nucleotide 20 and ends at 320 (Appendix). The polyadenylation signal begins at nucleotide 685 and the protein contains a 15 amino acid signal peptide. Three RR-1 mismatches (positions 69, 71, and 74), the absence of the two terminal prolines, and the immediate termination of the ORF following the consensus set this protein apart from the other RR-1



Figure 1. Eleven of the known RR-1-containing proteins from crustaceans as well as five from this study are shown in this alignment: CpCP/AM11.14 (P81575), CpAM11.59 (P81576), HaAMP1A (P81384), DD9A (BAA90875), DD9B (BAA90876), CpAM11.99 (P81577), CpAM12.39 (P81578), CpAM12.74 (P81579), HaAMP5 (P81389), CsAMP8.1 (AAV28467), and CsAMP6.0 (AAV28477). HaAMP1A is representing the group that includes HaAMP1B (P81385), HaAMP2 (P81386), HaAMP3 (P81387), and HaAMP4 (P81388). The five proteins from this study are underlined. The RR-1 and upstream consensus sequences are shown below the alignment. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids.

CpAM/CP11.14	(1)	-----DRDATILKDDRTDNG
CpAM11.59	(1)	-----ENTATVVVDERTDNG
HaAMP1A	(1)	-----DRDAQTLTDERSDQG
HaAMP5	(1)	-----QAPKIIETITDNREDDG
DD9A	(1)	-----VLLLVCLA AVAFARPDGDAELLLDEREDQG
DD9B	(1)	-----VLLLVCLA AVAFARPDGDARLLLDEREDQG
CpAM11.99	(1)	-----QFRPSVSRPDYKHIAIVSDNRYDNG
CpAM12.39	(1)	-----EHEAEIILDERQDNG
CpAM12.74	(1)	-----QLANEPPIEIIRQUESTDNG
CsCP14.1	(1)	-----APTYDSSSEKEAPILKYDRIQED
CsAMP9.3	(1)	-----RPQAPGDAQTLRSENINEG
CsAMP16.5	(1)	AKLPGHVDGGHDEHHQEHQEHHEHHDAGFVHSGDFRHEQPHVAILFDDRQAPA
CsAMP13.4	(1)	-----RPQSDPDHIEILRQETVDNG
CsAMP16.3	(1)	-----APQFLRTSRQPPIFPIPPQNNRILPPQNNRIIQVLRDSRDDRG
CsAMP8.1	(1)	-----RPDGD AETIVDERSDNG
CsAMP6.0	(1)	-----RPDKDATILTDERDRG
Upstream Consensus		<b>DRDA ILTDER DQG</b>

CpAM/CP11.14	(16)	DGNFHYSFETSNGIQDTKTGVP-GSAGQSNMNGDFSFP LDDGSTASFTYVADE
CpAM11.59	(16)	DGNFNYNFQTSNGIEDTKTGTP-GSQGQSNMQGTFRLLPDGTTAEVRYVADE
HaAMP1A	(16)	DGNFRYEFETSNGIYTQKTGTP-GSEGQSNYQGSFRFTLEDGTIAEVTYIADE
HaAMP5	(18)	AGNYFYE FETENGIRQSVRGTP-GAAGAVIKTGSFSFP LDDGTLAEFIFEADE
DD9A	(31)	DGNFRYVFETSNGIYQETVGTP-GASGQSNMVGSYRFT EPDGNVIEVRFTADE
DD9B	(31)	DGNFRYVFETSNGIFKETVGTP-GAEGQSNMVGSYRFTDPGGNVVEVRFTADE
CpAM11.99	(26)	DGNFGYDFETE HGINVEATGKP-GSKGQSNIGGSYRFILPDGTTAEVRYFADE
CpAM12.39	(16)	DGNFNRYRFETTNGIAEERVGV P-GSQGQSNMKGYSFNLPDGSRFQLSFAADE
CpAM12.74	(20)	DGNFNFLFETANGIYKEVSGYP-TANGAQAMTGSFRFP LDDGQIVEVSFTADE
CsCP14.1	(24)	DGRYNYEFETGNIGISHSQSGSPGDETCAINKAGHYTYTAPDGTIVVEMKFVANE
CsAMP9.3	(20)	NGVFQYNFETDNGIIVVEASGSP-NAEGSSDITGSYKFPLGNQFLEVTVYSDA
CsAMP16.5	(54)	DGSYATNFETEDGVRVSENGQP-GSAGQSNVEGSYSFTDPDGNLVEVRVVADE
CsAMP13.4	(21)	DGNFNLYLFETENGIYKEVVGSP-SANGAQAMRGSYRFP LDDGTIVEVTFVTADE
CsAMP16.3	(47)	DGNYNYEFETENGIYTNVEGRT-GLSGQTNQAGSFRFTLPDGT LAEVTFFVADE
CsAMP8.1	(18)	DGNFQYRFETSNGIIVEQRLGAP-GSEGQSNMQGDFGFTLPEGDRFDLTYVADE
CsAMP6.0	(18)	DGNFFYRFETSNGIQKEKTGTP-GSEGQSNMVGSGFQFP LDDGSTATFTFFVADE
Upstream		<b>DGNFRYEFETSNGIY KTGTP-GS QSNMQ</b>
RR-1		<b>G G Y A E</b>

CpAM/CP11.14	(68)	NGYHVESPLLPSIPE-----YVQKQIDFAAEQ RARGVIFD-----
CpAM11.59	(68)	FGYRPESPLL PVGPPELPP----HVHEL LRIAEEQRAQGITFE-----
HaAMP1A	(68)	NGFQPSSDLLPVGP PAPP----HVQRLLEIAAEQRAQGITFD-----
HaAMP5	(70)	YGYRVDSPLIPVAPPNPS----HVEELLQIVAQLKAQGAQWNDQGERID----
DD9A	(83)	NGFVPESDAIPQPPPLPA----HVEELLEIAEQQRDGRFTDQG GF EI-----
DD9B	(83)	NGFVPESDAIPQPPPLPA----HVEELLEIADQORREGIEFDDQGFPI-----
CpAM11.99	(78)	LG YRAESPLIPTPHPLPA----HAIEQIRFAESQR-----
CpAM12.39	(68)	NGYNADSPFIPTDHPLPA----HVIEL LALVEELKRQGATWDDKGVRIT----
CpAM12.74	(72)	NGYLPVSDFIPTPHPIPA----HVLET LAIVDELVRQGATWDEQGRIT----
CsCP14.1	(77)	NGFQPESPLL PVAP EFPHPPI PQFVLDQIAFAAEQDAARAGELSSEEGSYSYN
CsAMP9.3	(72)	RGFRPSTRYVSQRK-----
CsAMP16.5	(106)	FGFRAESPYVPTPHPLPA----HALQQIAYAEELRRLREQNGEPQFS-----
CsAMP13.4	(73)	NGYLPQSDAIP TPHPLPA----HVIET LALVDELVRQGATWDDQGRITRRK-
CsAMP16.3	(99)	AGFRAQSPLL PQAPPMPA----HALEHIRKAE EERARGVQFDSRGFRIGGF--
CsAMP8.1	(70)	NGYQPN SAFIPTDHPLPA----HVVELL AIVEELRRQGATWNDQGERLTRRK-
CsAMP6.0	(70)	NGYRVESPLLPIPE-----YVQRQIDFANSQ GKRR-----
RR-1		<b>GY P P</b>

arthrodial membrane proteins in this alignment (Figure 1). CsAMP9.3 is most similar to CsAMP8.1 (25%), HaAMP4, 3, and 2 (33%, 33%, and 32% identity, respectively). These homologous sequences differ from the RR-1 consensus in only one amino acid each. When compared to the upstream consensus, CsAMP9.3 shows 39% identity. CsAMP9.3 is also an acidic protein with a predicted isoelectric point of 4.93.

#### Calcifying Cuticle Proteins

The remaining four clones isolated from the EST library were shown to be expressed in calcifying hypodermis only (See Expression Data), and were named CsCP6.1, CsCP19.0, CsCP15.0, and CsCP14.1. CsCP6.1 (GenBank #DQ288153) is 416 bp and has an ORF of 74 amino acids from nucleotide 24 through 245 (Appendix). There is a polyadenylation signal at nucleotide 393. The first 20 amino acids of the virtual translation comprise a signal peptide. The RR consensus is also present with two mismatches at residues 64 and 67, and lacking the two terminal prolines. Since there is very little upstream sequence, this protein cannot be assigned a RR variant. CsCP6.1 was aligned with its most significant BLAST results (Figure 2). This protein is homologous to two proteins that also have very little sequence upstream from the RR consensus: *C. pagurus* cuticle protein 5.75 (59% identity) and *P. clarkii* calcification-associated peptide 2 (CAP-2) (36% identity). CsCP6.1 has a predicted isoelectric point of 4.82. Two transcripts in this study, CsCP15.0 and CsCP19.0, are very similar to each other. The complete sequence for CsCP15.0 (GenBank #DQ288154) is 621 bp with an ORF of 165 amino acids beginning at nucleotide 35 through 539 (Appendix). There is a polyadenylation signal at nucleotide 602. The entire CsCP19.0 sequence is 723 bp with an ORF of 182 amino acids starting at nucleotide 57 and ending at 647 (Appendix). The polyadenylation signal begins at

CsCP6.1	(1)	-----RPNEVLDFENDMTSHEQYGVPGTAVHGEYEAKD
CP5.75	(1)	-----GDIIDVDNDLFEHEQDGVAGTSVHGEYEAYD
CAP-2	(1)	MKLLVVVVLGLVVLAAARPSDIIDIEEDHLEHEQEGVPGTAVEGEYSWVA
RR-1		<b>G</b>
CsCP6.1	(34)	AFGNWYKVKYIADAGGFRVVS-----
CP5.75	(32)	AYGNEYEVKYIADHLGFRVL-----
CAP-2	(51)	PDGNEYKVKYVADHLGYRVLEDNVVPEVPELEDY
RR-1		<b>G Y A E G Y P P</b>

Figure 2. Alignment of virtually translated CsCP6.1 with its most homologous sequences in both Swiss-Prot and NCBI databases: *C. pagurus* CpCP5.75 (P81589) and *P. clarkii* CAP-2 (BAD16776). The RR-1 consensus is shown below the alignment in bold. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids.

nucleotide 707. CsCP19.0 was found to have two isoforms, a and b (GenBank #DQ288149 and DQ288150, respectively). These isoforms differ by only one nucleotide at position 543 where adenine in CsCP19.0a is replaced by guanine in CsCP19.0b. This changes the amino acid from isoleucine in CsCP19.0a to valine in CsCP19.0b. The first 15 amino acids of both CsCP15.0 and CsCP19.0 (a and b) comprise a signal peptide.

These proteins were aligned with each other and their most significant BLAST results (Figure 3). CsCP15.0 is most homologous to CpCP14.99 with 75% amino acid identity, whereas CsCP19.0 is only 49% identical. CsCP19.0 is most homologous to CpCP18.75 with 50% identity. CsCP15.0 is only 44% identical to this transcript. CsCP15.0 and CsCP19.0 are 50% identical to each other. CsCP19.0 is 42.1% identical to HaCP20.2 (a,b) and CsCP15.0 is only 39.8% identical. HaCP20.2 has two isoforms that differ at one residue (184) where alanine in (a) is replaced by proline in (b). All five of these proteins contain three copies of the insect-post18 motif and two of these copies are part of a larger 45 residue repeated region (Figure 3). CsCP15.0 has a predicted isoelectric point of 6.35 and CsCP19.0 (a and b) is more basic at 9.89. The final calcified cuticle transcript is CsCP14.1 (GenBank #DQ288151). It is 576 bp with an ORF of 129 amino acids (Appendix). The start codon begins at nucleotide 18 with the stop codon at 450. There is a polyadenylation signal at nucleotide 558. The first 15 amino acids of the virtually translated protein include a signal peptide. CsCP14.1 was put into the same alignment with the other RR-1 containing proteins (Figure 1). It shows 35% identity with CpAM11.59, CpAM/CP11.14, and HaAMP1A. The *Schistocera gregaria* endocuticular protein, SgAbd-2 (NCBI protein: Q7M4F3), has only 29% identity with CsCP14.1 and was not included in the alignment because it is an insect protein. CsCP14.1 contains RR-1 with only two

CsCP15.0 (1) LPVLPDDPLVAAEKAKFFFAAYQAALPATPPKPADP-----  
 CpCP14.99 (1) --TIQDAPDVAAEKARFFFLAYQAALPATPPKPADP-----  
 CsCP19.0A (1) MPFIIPDAPDVAAEKARFFQAYQVAHAANLPKASRPSHQAFVPHQAFNQAF  
 CpCP18.76 (1) --LIPDDPDVAAEKARFFRTFKIIEGASKPRGGGIIVRPALPPGADVITM  
 HaCP20.2A (1) --FIPDNPDVAAEKARFFQAFKAAEAAALPRFSPHQSHDTGHYTPTRYG  
 Insect-post **V DTPEVAAAKAAF AAY**  
 18 motif #1

CsCP15.0 (36) -----PKWYGPLASSVPAGLPGSSPVVAPTAEVAAARNEF  
 CpCP14.99 (34) -----PKWYGPLASKIPAGLPGSSANVALTADLSAARNAF  
 CsCP19.0A (51) VQT-----PKWMGPLASNVPAGLPGSSAFVADTPEVQAAKNHF  
 CpCP18.76 (49) PRPQ-----PKWMGPLASKVPASLPGSTAFVSETSDVQNARSHF  
 HaCP20.2A (49) TSHYNPPQYTHTPAQPKWTGPVAATVPAGVDGTLTPVSDTRDVAAARNTF  
 -----  
 Insect-post **V DTPEVAAAKAAF**  
 18 motif #2

CsCP15.0 (71) YSTYNAQVAAVAPT-----AGGPVFKVVPVAVT--GTWTGPVAATI  
 CpCP14.99 (69) YNTYNAQVAAVAPK-----SDGPVFVAVVPVATS--AQWTGPVAATL  
 CsCP19.0A (89) LNAISAQVAATVPVGPRTSQPSFAPRPVFQASAPAPRPQPKWTGPLASQV  
 CpCP18.76 (88) FNTYNAQVAATMPSPDSPTYYSAPSAPAYVPDAPQEK----WTGPLASAV  
 HaCP20.2A (99) FNTYRSQVAATLGAAPRPHG-GAVQYQAAPSYRPAPSYQPKWTGPLASNV  
 -----  
 Insect-post **AAAY**  
 18 motif #2

CsCP15.0 (110) PAGLPGSAPNVADTADVAGAKTAFEDTYQRQVAAVAPAPKV---  
 CpCP14.99 (108) PAGVPGSSPNVAETADVLNAKNAFFETTYNQQVAAVAPAPKV---  
 CsCP19.0A (139) PAGLPGSSPILSDTAEVAAAKNAFFHTYSAQVAATAGAPSTRFF  
 CpCP18.76 (134) PAGLPGSSPVVFDTPPEVYNAAKNAFFNTYKNQVKAIIPRPSYF--  
 HaCP20.2A (148) PAGLPGSTNQVADTPEVAAAKHSFYNTYQRQAAAAAPSRY--  
 -----  
 Insect-post **V DTPEVAAAKAAF AAY**  
 18 motif #3

Figure 3. Alignment of virtually translated CsCP15.0 and CsAMP19.0 (a shown) with their most homologous sequences in both Swiss-Prot and NCBI databases: *C. pagurus* CpCP14.99 (P81583), CpCP18.76 (P81584) and HaCP20.2 (a shown) (Q7M496). The post-molt 18 amino acid motif is shown and the repeated regions are denoted by dashes below the alignment. NCBI protein database numbers are given after each protein. Black highlight indicates identical amino acids, dark gray indicates conserved amino acids, and light gray indicates a block of similar amino acids.

mismatches at residues 72 and 79 where tyrosine is replaced by phenylalanine. The N-terminal region is 37% identical to the upstream consensus. CsCP14.1 has a predicted isoelectric point of 4.31.

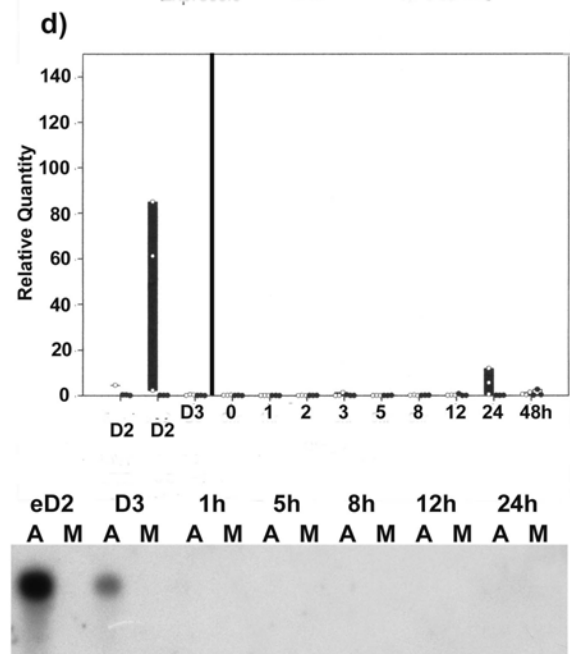
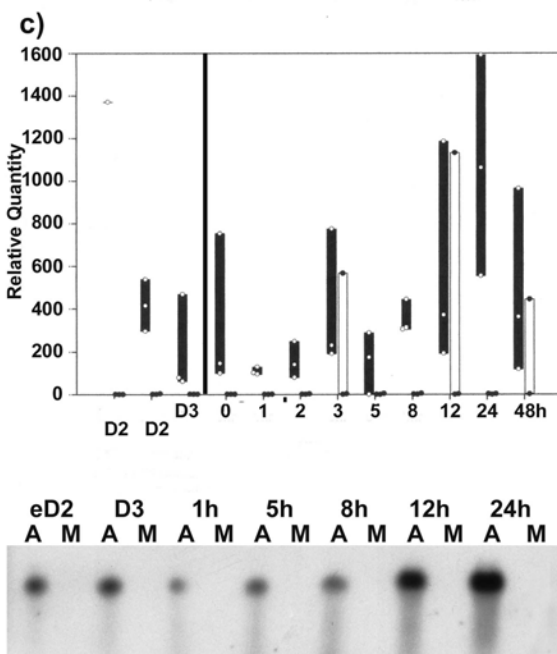
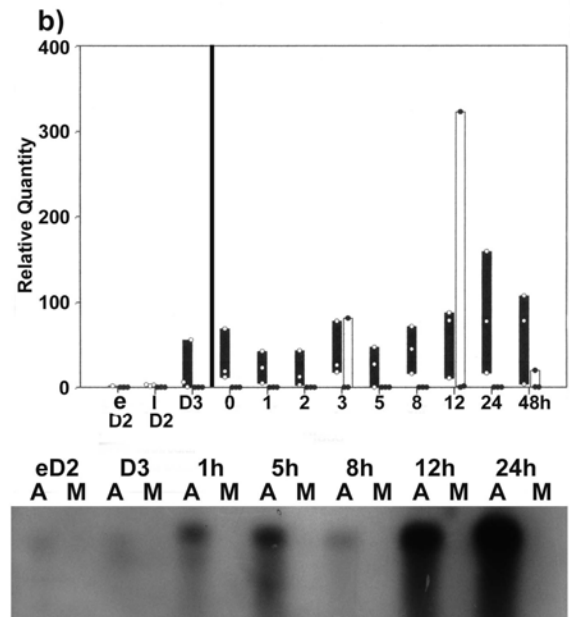
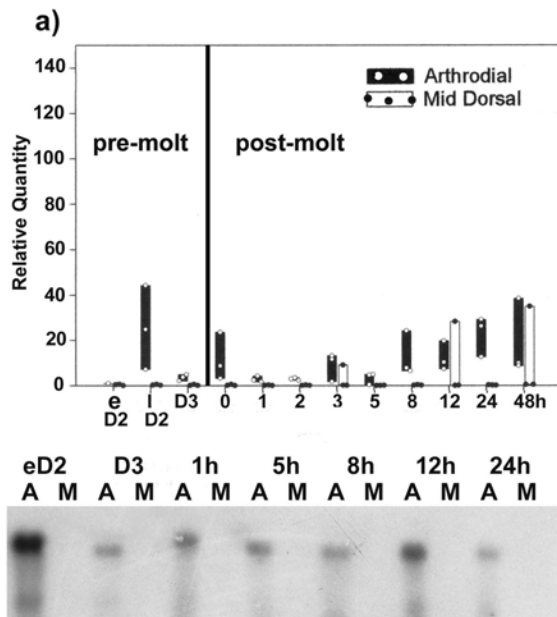
#### Expression Data

Northern blotting and qPCR were used to determine in which tissue and at what time points each transcript is expressed. For qPCR, cDNAs from three pre-molt stages (early D<sub>2</sub> (e-D<sub>2</sub>), late D<sub>2</sub> (l-D<sub>2</sub>), and D<sub>3</sub>) and eight post-molt time points (0, 1, 3, 5, 8, 12, 24, and 48 h) were used. Only one sample of e-D<sub>2</sub> arthrodial cDNA was quantified for each transcript due to a lack of RNA whereas all others were analyzed in triplicate. For Northern blots, RNAs from pre-molt stages e-D<sub>2</sub> and D<sub>3</sub>, and post-molt time-points 1, 5, 8, 12 and 24 h from both arthrodial membrane and mid-dorsal tissue were used to determine these expression patterns.

Four of the eight transcripts were expressed primarily in arthrodial membrane hypodermis. Three of these four (CsAMP16.5, CsAMP16.3, and CsAMP13.4) showed expression at all time points (Figure 4). CsAMP16.5 showed a maximum expression of approximately 45 during l-D<sub>2</sub> and a minimum near 0 immediately following ecdysis (Figure 4a). CsAMP16.3 expression during the pre-molt stages was very low but detectable (most noticeably in the Northern blot) and had a maximum of approximately 150 at 24 h post-ecdysis (Figure 4b). CsAMP13.4 showed the highest expression of all arthrodial transcripts and was maximal 24 h post-ecdysis (Figure 4c). The occasional expression in mid-dorsal tissue (3, 12, and 48 h) was measured in the same three crabs for each transcript. The fourth transcript, CsAMP9.3, was expressed exclusively in arthrodial membrane hypodermis, but only during pre-molt stages (Figure 4d). The Northern blot showed the highest expression during pre-molt stage e-D<sub>2</sub>; its expression was continued during l-D<sub>2</sub> as seen in the qPCR graph and was lowest during D<sub>3</sub> in the Northern blot.

Figure 4. qPCR graphs and Northern blots for (a) CsAMP16.5, (b) CsAMP16.3, (c) CsAMP13.4, and (d) CsAMP9.3. A vertical bar separates pre-molt stages from post-molt time points. Early and late D<sub>2</sub> stages are denoted eD2 and lD2. Data for individual crabs is indicated by the open or closed circles and the bars indicate the range, or variation, in the data. (Note: The scales are not the same on each graph and only one sample from the arthroal e-D<sub>2</sub> stage was measured.)





The other four transcripts in this study were expressed primarily in mid-dorsal hypodermis. Two of these transcripts (CsCP6.1 and CsCP19.0) were expressed only after ecdysis (Figure 5 a, b). CsCP6.1 showed maximal expression of 400 at 24 h after ecdysis. Expression appeared to increase to that point and may be much lower 48 h post-ecdysis (Figure 5a). Both isoforms of CsCP19.0 could have been detected in the Northern blot and qPCR analysis since a unique probe and primer set could not be designed for each isoform. This transcript had an expression maximum of 140 at 24 h post-ecdysis, but seemed to have a rather stable expression level prior to that (Figure 5b). Expression seemed to be lower 48 h post-ecdysis just as in CsCP6.1. CsCP15.0 had the highest expression levels of this group, ranging from 0 to 1600, and was present both before and after ecdysis. Its expression increased gradually, remained high between 1 and 24 h post-molt, and was much lower at 48 h (Figure 5c). Occasional expression of these three transcripts was seen in the post-molt arthrodial samples. The final transcript, CsCP14.1, did not show any arthrodial expression and was only present during the pre-molt stages in mid-dorsal hypodermis (Figure 5d).

#### *In situ* Hybridization

*In situ* hybridization was performed to confirm the expression data and also to determine whether or not transcripts were expressed uniformly in their specific tissue. Tissue types were differentiated by staining with hematoxylin and eosin, which preferentially stains the basophilic endocuticle layer that will be calcified and not the arthrodial membrane cuticle (Williams et al., 2003) (Figure 6a). The hypodermis can therefore be identified as depositing either arthrodial membrane or calcifying cuticle. Figures 6a and 6b show the same pre-molt (D<sub>2</sub>) tissue section, but the tissue in 6b was not hybridized with any probe. There was no AP-staining in any tissue type. Sense probes were used to test for nonspecific probe binding (Figure 6c) and some

Figure 5. qPCR graphs and Northern blots for (a) CsCP6.1, (b) CsCP19.0, (c) CsCP15.0, and (d) CsCP14.1. A vertical bar separates pre-molt stages from post-molt time points. Early and late D<sub>2</sub> stages are denoted eD<sub>2</sub> and lD<sub>2</sub>. Data for individual crabs is indicated by the open or closed circles and the bars indicate the range, or variation, in the data. (Note: The scales are not the same on each graph and only one sample from the arthroal e-D<sub>2</sub> stage was measured.)

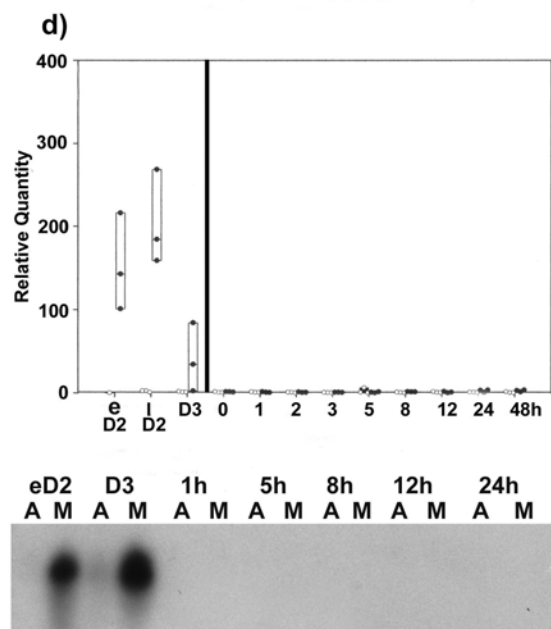
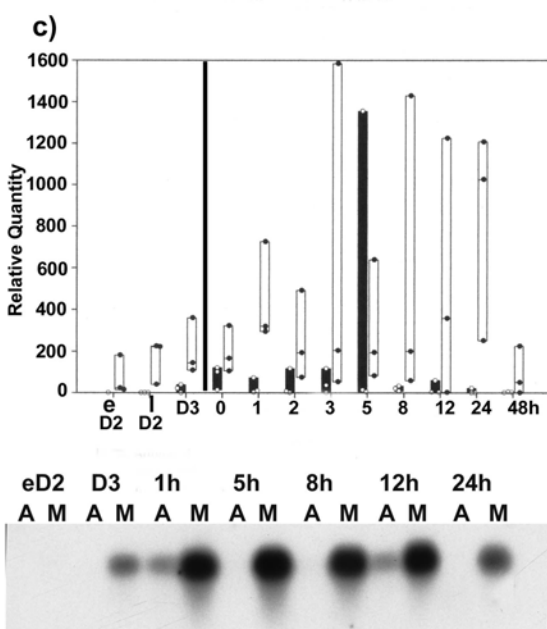
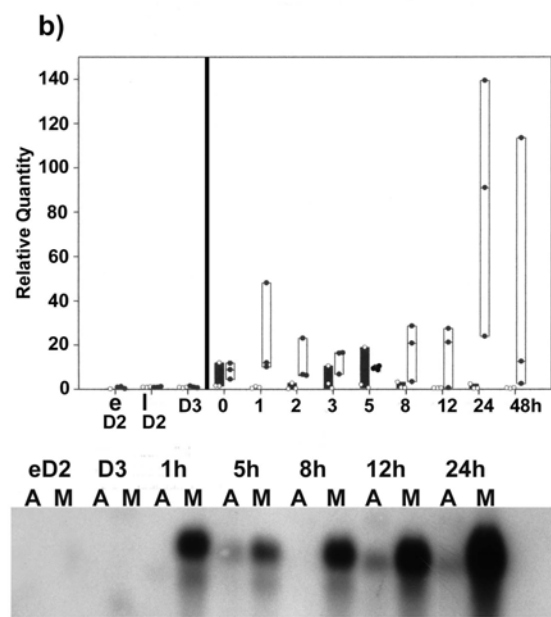
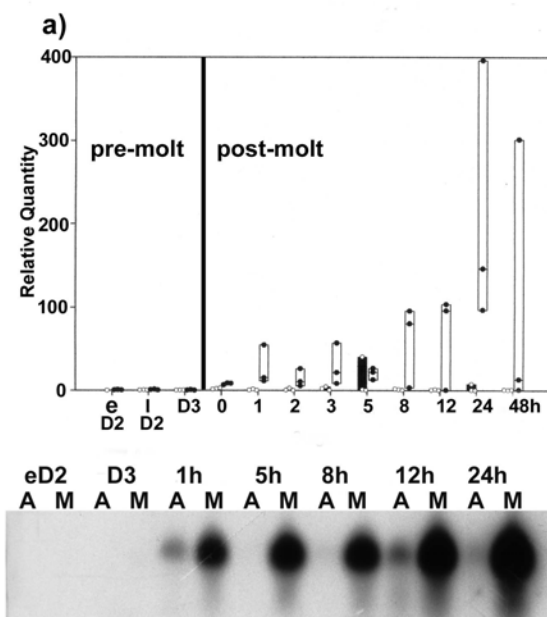
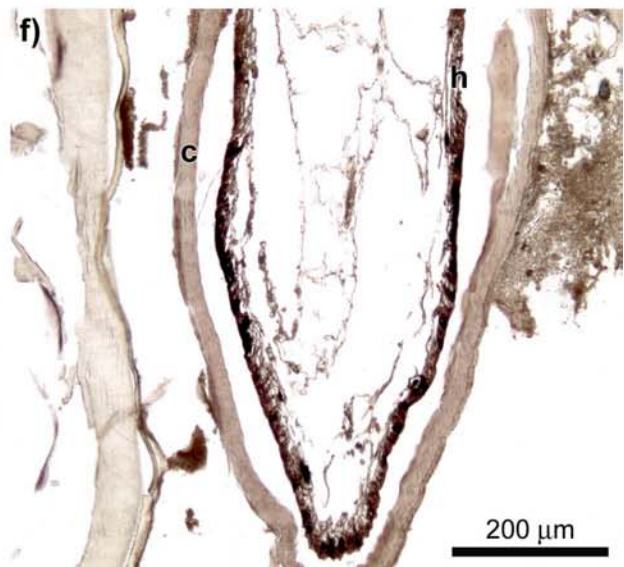
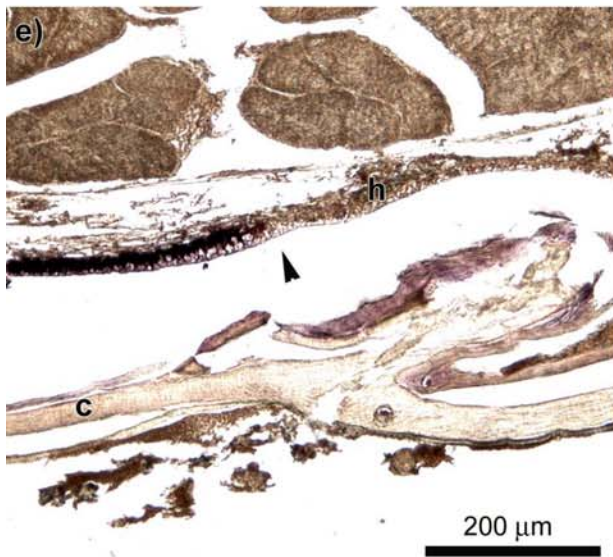
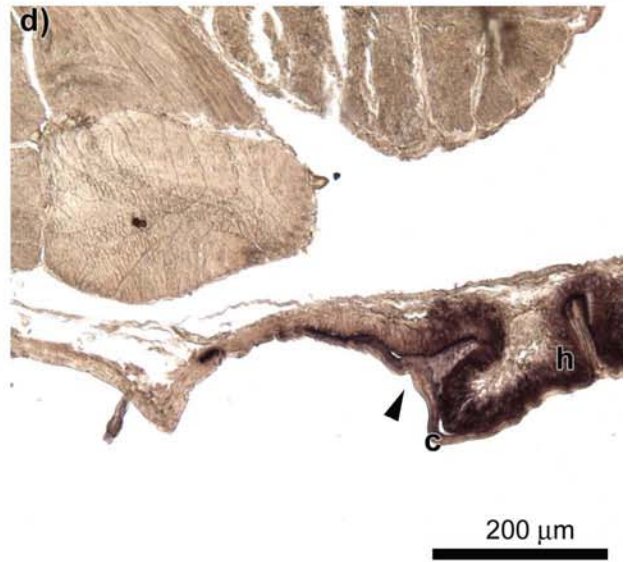
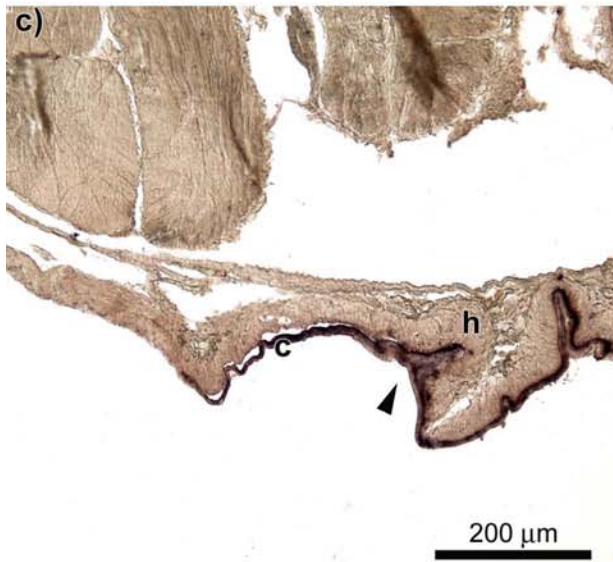
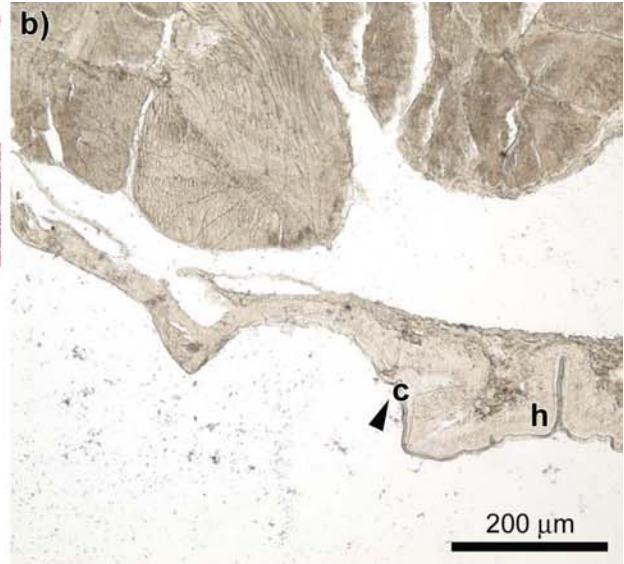
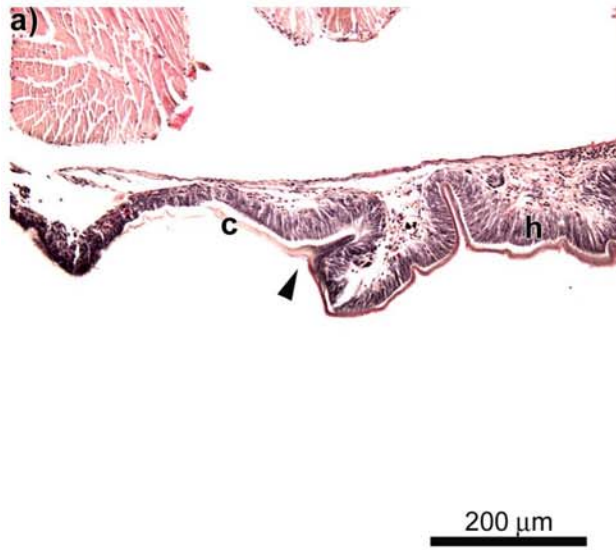


Figure 6. Sections of the walking leg joint from pre-molt stage D<sub>2</sub> (a-d), 48 h post-molt (e), and 12 h post-molt (f) crabs. Only arthroal hypodermis is shown in (f). Sections were (a) stained with hematoxylin and eosin, (b) used as a control in *in situ* hybridization in which no probe was hybridized, (c) hybridized with the sense probe for CsCP14.1, (d) hybridized with the anti-sense probe for CsCP14.1, (e) CsAMP16.5, and (f) CsAMP16.3. In sections a-e, hypodermis to the right of the arrowhead synthesizes calcifying cuticle and to the left synthesizes arthroal membrane cuticle (non-calcifying). Both hypodermis (h) and cuticle (c) are labeled on each section.



background staining was present in all tissues. In addition, the calcifying cuticle in post-molt sections as well as the calcified exuvium present in pre-molt sections seemed to have a high affinity for the nucleotides of the probe. Nonspecific cuticle staining was also seen in *M. japonicus in situ* hybridization experiments (Watanabe et al., 2000; Ikeya et al., 2001; Endo et al., 2004). When anti-sense probes were used, all the transcripts were detected in the tissue type expected from their expression data (Figure 6d, e, f). Only CsAMP16.3 and CsCP15.0 were not expressed uniformly in the hypodermis. In all sections, the separation of the cuticle from the hypodermis can be seen and is an artifact of fixation not affecting the results.

## DISCUSSION

This study has determined the sequence and expression patterns for eight transcripts in *C. sapidus* (Table 2). All the translated proteins contained a signal sequence indicating that they are exported from the hypodermis. Four of these are similar in sequence, contain the RR-1 consensus (Figure 1), and are expressed in non-calcifying arthrodial hypodermis (Figure 4). CsAMP9.3 has the most degenerate RR-1 consensus and is not as similar to its most significant BLAST results as are the other arthrodial membrane proteins. This is because its protein sequence terminates immediately following the RR-1 consensus. CsAMP9.3 was expressed primarily during the pre-molt stages indicating that it is a constituent of the epi- and exo-cuticle layers (Figure 4d). Its extremely low expression level suggests that it does not play a structural role in the arthrodial membrane. CsAMP16.3 is expressed from late pre-molt through 48 h post-molt increasing only slightly to 24 h so, it would be present in all cuticle layers except the epicuticle (Figure 4b). CsAMP13.4 was most highly expressed 24-48 h post-molt indicating that it is an endocuticle protein (Figure 4c). It is very similar to two *C. pagurus* arthrodial membrane proteins (CpAM12.74 and 12.39). Preliminary *in situ* data have shown that CsAMP13.4 is

Table 2

Summary of expression and sequence data with proposed functions for all eight transcripts

TRANSCRIPT	LOCATION	EXPRESSION	MOTIF	PROPOSED FUNCTION
CsAMP9.3	arthrodial	Pre-molt	RR-1	functional protein in epi- and exocuticle layers
CsAMP16.5	arthrodial	Pre- and Postmolt	RR-1	structural protein in all cuticle layers
CsAMP16.3	arthrodial	Pre- and Postmolt	RR-1 and RGD	anchors pore canals to chitin fibers
CsAMP13.4	arthrodial	Pre- and Postmolt	RR-1	structural protein in all cuticle layers
CsCP14.1	calcifying	Pre-molt	RR-1	inhibits calcification in epi- and exocuticle layers
CsCP6.1	calcifying	Postmolt	partial RR	promotes calcification in endocuticle
CsCP15.0	calcifying	Pre- and Postmolt	insect-post18	protein-protein or protein-chitin interactions
CsCP19.0a/b	calcifying	Postmolt	insect-post18	protein-protein or protein-chitin interactions



currently the only arthrodial transcript found in the non-calcifying inner branchial chamber hypodermis. CsAMP16.5 decreases in expression from e-D2 to ecdysis, remains low during the 1-3 h PECA (Shafer et al., 1995), and increases through 48 h post-molt (Figure 4a). CsAMP16.5 is homologous to the *C. pagurus* arthrodial membrane proteins CpAM11.99 and 12.99. This data suggests that CsAMP16.5 and CsAMP16.3 could be structural elements common to all arthrodial membrane cuticles while CsAMP13.4 could be common to all non-calcifying cuticles.

CsAMP9.3 and CsAMP13.4 both end in the basic residues RK (RRK in CsAMP13.4), as do two other *C. sapidus* arthrodial membrane proteins, CsAMP6.0 and CsAMP8.1 (Wynn and Shafer, 2005), and several calcified cuticle proteins ending with RxKR (Kennedy, 2004). The C-terminal RK residues were also seen in the virtually translated cDNA of CAP-1, but not in the extracted protein (Inoue et al., 2003). These basic residues are not present in the *H. americanus* or *C. pagurus* extracted proteins and it is assumed that they are removed from the mature *C. sapidus* proteins. The significance of the removal is unknown and caused no decrease in calcification activity of CAP-1 (Inoue et al., 2003).

CsAMP16.3 has a RGD domain (Appendix), also found in CsAMP6.0 (Wynn and Shafer, 2005), that is the key sequence in a cellular attachment domain found in extracellular matrix proteins (Pierschbacher et al., 1984; Yamada, 1991). This region could play a role in the connection between cuticle and hypodermis, and was experimentally shown to have cell-adhesive activity in *Bombyx mori* cuticle protein 30 (BMCP30) (Nakato, 1993). CsAMP16.3 could be involved in anchoring the pore canals to the cuticle by providing a connection between the chitin fibers and hypodermal cell membranes. These pore canals do not extend into the epicuticle (Roer and Dillaman, 1984), which would explain the lack of transcription of this protein during epicuticle deposition (Figure 4b). A fiber (possibly chitin, or chitin-protein) has

recently been found to be intimately associated with the pore canals (Modla and Dillaman, personal communication) supporting this hypothesis. This function would be unique and not shared by the proteins most similar to CsAMP16.3 because they do not contain the RGD domain. In fact, the previously described RGD-containing CsAMP6.0 is very different from CsAMP16.3, but their expression patterns are similar and they could play similar roles in the cuticle.

All of these arthrodial membrane proteins contain the RR-1 consensus that has been experimentally shown to bind chitin (Rebers and Willis, 2001). To better understand how this region associates with chitin, 8 hard and 19 soft cuticle RR-containing insect proteins were analyzed for secondary structures by six different algorithms (Iconomidou et al., 1999). All six algorithms predicted the formation of  $\beta$ -pleated sheets with the three invariant glycine residues of the basic RR consensus corresponding exactly at maxima of  $\beta$ -turn/loop prediction. Hard cuticle proteins were found to have 4  $\beta$ -sheets while soft cuticles contained 3. This  $\beta$ -pleated sheet structure was also predicted using homology modeling with an insect soft cuticle protein (Hamodrakas et al., 2002). For quite some time, the  $\beta$ -pleated sheet has been thought to define the helicoidal structure of the cuticle by its association with chitin (Hamodrakas et al., 2002; Rebers and Willis, 2001; Iconomidou et al., 1999). The aromatic residues phenylalanine and tyrosine are frequently found within the RR consensus, belong to the same hydrophobic face of the  $\beta$ -sheets, and could stack against the faces of the saccharide rings of chitin (Hamodrakas et al., 2002). A two turn  $\alpha$ -helix starting and ending by the two terminal prolines of the RR consensus was predicted by homology modeling (Hamodrakas et al., 2002), but not by the secondary structure algorithms (Iconomidou et al., 1999). This feature is absent in CsAMP9.3 since it lacks the two terminal prolines (Appendix). A function for this region has not been proposed so the displacement of one or both prolines by one residue (CsAMP13.4 and

CsAMP16.5) is unknown. The proposed model adopts an ideal shape to cover chitin fibers with the aromatic ring face while the outer face promotes regular 3-D packing of chitin for the formation of the helicoidal structure of the cuticle (Hamodrakas et al., 2002). Experiments with both the RR-1 (Togawa et al., 2004) and RR-2 (Rebers and Willis, 2001) containing proteins revealed that the N-terminal regions are important for chitin binding and not just the consensus itself. It has been suggested that the different RR variants interact with chitin fibers in subtly different ways (Togawa et al., 2004). The basic consensus is still important, however, and the removal of the two conserved tyrosine residues resulted in a loss of chitin-binding ability (Rebers and Willis, 2001).

Four transcripts expressed in this study were expressed in hypodermis that secretes calcifying cuticle (Table 2). The complete sequences for these transcripts were determined (Appendix) as well as their expression patterns (Figure 5). Two of these proteins (CsCP14.1 and CsCP6.1) contain a RR consensus. CsCP14.1 contains RR-1 and was expressed only during pre-molt stages (Figure 5d). The abrupt disappearance of this transcript after ecdysis suggests that it may either be restricted to epi- and exocuticle deposition or involved in inhibiting calcification in these pre-exuvial layers. The latter would support the model proposed by Coblentz et al. (1998). This protein is most similar to two arthroal membrane proteins from *C. pagurus* (CpAM11.59) and *H. americanus* (HaAMP1A). It is also similar to two proteins found in both tissue types, CpAM/CP11.14 and SgAbd-2. CpAM/CP11.14 was extracted from calcifying and non-calcifying cuticle from *C. pagurus* (Andersen, 1999) and SgAbd-2 was extracted from both hard (thorax) and soft (abdominal membrane) cuticle regions of the desert locust, *Schistocera gregaria* (Andersen, 1998b). Although these four proteins are the most significant BLAST results for CsCP14.1, they are much more similar to each other than they are to CsCP14.1. This

protein is thought to be involved in the inhibition of nucleation in the cuticle, but it cannot be determined by comparison to another protein with this function because one has not yet been found. It is distinct from the other RR-containing proteins and shows little similarity in an alignment (Shafer and Faircloth, 2006). CsCP14.1 could be the molecule inhibiting two previously described *C. sapidus* proteins (CsCP8.5 and CsCP8.2) expressed during the pre-molt stages and thought to be involved in  $\text{CaCO}_3$  nucleation (Wynn and Shafer, 2005). It has been suggested that inhibition of these proteins prevents calcification in the pre-exuvial layers until after ecdysis.

The other RR-containing calcified cuticle protein, CsCP6.1, has a RR consensus that is missing the two terminal prolines and cannot be assigned a RR variant because of its short N-terminal region (Figure 2). This protein is most similar to CAP-2 and CpCP5.75, which cannot be assigned a RR variant for the same reason. CAP-2 was experimentally found to have one  $\text{Ca}^{2+}$ -binding site as compared to CAP-1 which has two (Inoue et al., 2004). CsCP6.1, CpCP5.75 and CAP-2 have a similar N-terminal region (HEQxGVPGTAV) that is not present in CAP-1. This region could be the  $\text{Ca}^{2+}$ -binding site since there is very little sequence after the RR consensus. CsCP6.1 is expressed at maximal levels during endocuticle deposition and calcification (Figure 5a), and could play a role in  $\text{CaCO}_3$  nucleation in the endocuticle.

There are two other calcified cuticle proteins that do not contain a RR consensus, CsCP15.0 and CsCP19.0 (a and b). These proteins are very similar to each other, to two *C. pagurus* calcified cuticle proteins (CpCP14.99 and CpCP18.76), and to one *H. americanus* calcified cuticle protein (HaCP20.2) that also has two isoforms, a and b (Figure 3). These five proteins each contain three copies of the insect-post18 motif (Andersen, 1999, 2000). So far it has only been found in proteins from post-molt cuticles, but CsCP15.0 shows some expression

during the pre-molt stages (Figure 5c). CsCP19.0 is only expressed post-molt with maximum expression 24 h after ecdysis (Figure 5b). CsCP15.0 and CpCP14.99 share the most identical amino acids in the N-terminal copy of the insect-post18 motif (61.1% and 55.6%, respectively). CsCP19.0 is 66.7% identical to the central copy while CpCP18.76 and HaCP20.2 (a and b) are 66.7% identical to the C-terminal copy. The motifs in the central and C-terminal regions are actually part of a larger 46 residue repeat from position 63-108 and 136-181 in HaCP20.2 (Figure 3) (Nousiainen, 1998). This larger repeat is also present in the other four proteins. The only other crustacean proteins with the insect-post18 motif are HaCP18.8 and HaCP14.2 (Andersen, 2000). They contain a single copy of the motif at the C-terminus and also contain the only RR-3 variants found in crustaceans. HaCP18.8 has the higher sequence conservation in these regions while HaCP14.2 has the least in both the insect-post18 motif and RR-3 (Andersen, 2000). The insect-post18 motif has been suggested to play a role in protein-protein or protein-chitin interactions (Nakato et al., 1990).

*In situ* hybridizations confirmed the expression patterns seen in the Northern blots and qPCR data (Figure 6). Two transcripts, CsCP15.0 and CsAMP16.3, were the only ones not expressed uniformly throughout the hypodermis. CsAMP16.3 was expressed in “spots” throughout the arthrodial membrane hypodermis (Figure 6f). At 3 hr post-molt, CsCP15.0 was expressed more at the boundary between arthrodial and calcified cuticle and by 12 hr post-molt was uniformly expressed throughout calcifying cuticle hypodermis. In this region it could serve a structural role thus strengthening the transition area from calcified to arthrodial cuticle as soon after the molt as possible and then spreads to the rest of the calcifying hypodermis. Since this transcript is expressed at very high levels and maximally during endocuticle deposition, it could serve a function in the endocuticle of calcified cuticle throughout the organism.

The functions of the transcripts discussed are still speculative and further study using RNA-interference and immunohistochemistry techniques will be useful in testing the hypotheses put forward here.

## LITERATURE CITED

- Addadi, L., Weiner, S., 1985. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. *Proc. Natl. Acad. Sci. USA* 82, 4110-4114.
- Andersen, S.O., 1998a. Characterization of proteins from arthrodial membranes of the lobster, *Homarus americanus*. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 121, 375-383.
- Andersen, S.O., 1998b. Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* 28, 421-434.
- Andersen, S.O., 1999. Exoskeletal proteins from the crab, *Cancer pagurus*. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 123, 203-211.
- Andersen, S.O., 2000. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochem. Mol. Biol.* 30, 569-577.
- Coblentz, F.E., Shafer, T.H., Roer, R.D., 1998. Cuticular proteins from the blue crab alter in vitro calcium carbonate mineralization. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 121, 349-360.
- Coblentz, F.E., Towle, D.W., Shafer, T.H., 2006. Expressed sequence tags from normalized cDNA libraries prepared from gill and hypodermal tissues of the blue crab, *Callinectes sapidus*. *Comp. Biochem. Physiol. D* (in press; on-line pre-publication).
- Dillaman, R.M., Hequembourg, S., Gay, D.M., 2005. Early pattern of calcification in the dorsal carapace of the blue crab, *Callinectes sapidus*. *J. Morphol.* 263, 356-374.
- Drach, P., Tchernigovtzeff, C., 1967. Sur la méthode de détermination des stades d'intermue et sone application gènèrale aux crustacés. *Vie Milieu, Ser. A Biol. Mar.* 18, 595-610.
- Elliott, E.A., Dillaman, R.M., 1999. Formation of the inner branchiostegal cuticle of the blue crab, *Callinectes sapidus*. *J. Morphol.* 240, 267-281.
- Endo, H., Persson, P., Watanabe, T., 2000. Molecular cloning of the crustacean DD4 cDNA encoding a Ca<sup>+2</sup>-binding protein. *Biochem. Biophys. Res. Commun.* 276, 286-291.
- Endo, H., Takagi, Y., Ozaki, N., Kogure, T., Watanabe, T., 2004. A crustacean Ca<sup>+2</sup>-binding protein with a glutamate-rich sequence promotes CaCO<sub>3</sub> crystallization. *Biochem. J.* 384, 159-167.
- Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., 2002. A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. Mol. Biol.* 32, 1577-1583.
- Horst, M.N., 1990. Concurrent protein synthesis is required for in vivo chitin synthesis in postmolt blue crabs. *J. Exp. Zool.* 256, 242-254.
- Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 1999. Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? *Insect Biochem. Mol. Biol.* 29, 285-292.
- Ikeya, T., Persson, P., Kono, M., Watanabe, T., 2001. The DD5 gene of the decapod crustacean *Penaeus japonicus* encodes a putative exoskeletal protein with a novel tandem repeat structure. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 128, 379-388.

- Inoue, H., Ozaki, N., Nagasawa, H., 2001. Purification and structural determination of a phosphorylated peptide with anti-calcification and chitin-binding activities in the exoskeleton of the crayfish, *Procambarus clarkii*. *Biosci. Biotechnol. Biochem.* 65, 1840-1848.
- Inoue, H., Ohira, T., Ozaki, N., Nagasawa, H., 2003. Cloning and expression of a cDNA encoding a matrix peptide associated with calcification in the exoskeleton of the crayfish. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 136, 755-765.
- Inoue, H., Ohira, T., Ozaki, N., Nagasawa, H., 2004. A novel calcium-binding peptide from the cuticle of the crayfish, *Procambarus clarkii*. *Biochem. Biophys. Res. Commun.* 318, 649-654.
- Kennedy, P.J., 2004. Characterization of a gene family associated with calcified structures in the blue crab, *Callinectes sapidus*. M.S. Thesis, University of North Carolina Wilmington, Wilmington, USA.
- Kragh, M., Molbak, L., Andersen, S.O., 1997. Cuticular proteins from the lobster, *Homarus americanus*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 118, 147-154.
- Nakato, H., Yoriyama, M., Izumi, S., Tomino, S., 1990. Structure and expression of a mRNA for a pupal cuticle protein of the silkworm, *Bombyx mori*. *Insect Biochem.* 20, 667-678.
- Nakato, H., 1993. Structure and expression of cuticle protein genes in silkworm, *Bombyx mori*. Ph.D. Dissertation, Tokyo Metropolitan University, Tokyo, Japan.
- Neville, A.C., 1975. *Biology of the arthropod cuticle*. Springer, New York, NY.
- Nousiainen, M., Rafn, K., Skou, L., Roepstorff, P., Andersen, S.O., 1998. Characterization of exoskeletal proteins from the American lobster, *Homarus americanus*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 119, 189-199.
- Pierschbacher, M.D., Ruoslahti, E., 1984. Variants of the cell recognition site of fibronectin that retain attachment-promoting activity. *Proc. Natl. Acad. Sci. USA* 81, 5985-5988.
- Rebers, J.E., Riddiford, L.M., 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *J. Mol. Biol.* 203, 411-423.
- Rebers, J.E., Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol.* 31, 1083-1093.
- Roer, R.D., Dillaman, R.M., 1984. The structure and calcification of the crustacean cuticle. *Amer. Zool.* 24, 893-909.
- Roer, R.D., Dillaman, R.M., 1993. Molt-related change in integumental structure and function. In: Horst, M.N., Freeman, J.A. (Eds.), *The Crustacean Integument*, CRC Press, Inc., Boca Raton, pp. 1-37.
- Shafer, T.H., Roer, R.D., Midgette-Luther, C., Brookins, T.A., 1995. Postecdysial cuticle alteration in the blue crab, *Callinectes sapidus*. *J. Exp. Zool.* 271, 171-182.
- Shafer, T.H., Faircloth, L.M., 2006. Identifying exoskeleton proteins in the blue crab from an expressed sequence tag (EST) library. *Integr. Comp. Biol.* In Review.
- Togawa, T., Nakato, H., Izumi, S., 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 34, 1059-1067.



- Watanabe, T., Persson, P., Endo, H., Kono, M., 2000. Molecular analysis of two genes, DD9A and B, which are expressed during the postmolt stage in the decapod crustacean *Penaeus japonicus*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 125, 127-136.
- Williams, C.L., Dillaman, R.M., Elliott, E.A., Gay, D.M., 2003. Formation of the arthroal membrane in the blue crab, *Callinectes sapidus*. *J. Morphol.* 256, 260-269.
- Wynn, A., Shafer, T.H., 2005. Four differentially expressed cDNAs in *Callinectes sapidus* containing the Rebers-Riddiford consensus sequence. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 141, 294-306.
- Yamada, K.M., 1991. Adhesive recognition sequences. *J. Biol. Chem.* 266, 12809-12812.

## APPENDIX

Appendix. Nucleotide sequences for each transcript with the virtual translations below. Signal sequences are underlined and polyadenylation signals are enclosed by a box. Regions of import are highlighted where applicable.

### CsAMP16.5

```

1  ccgcagatatgaagctcgctcggttcgctgccttggtcgggcggtgtgtcttggggccaagc 60
-15      M K L V V F A A L V G V C L G A K 2

61  tgcctgggtcacgtggatggcgccacgacgaacaccaccaagagcaccaggagcaccacg 120
3  L P G H V D G G H D E H H Q E H Q E H H 23

121 agcaccacgacgctggcttcgtacacagcggagatttccgccatgaacagcctcacgtag 180
24 E H H D A G F V H S G D F R H E Q P H V 43

181 ccatacctgttcgacgacccgccaggccccagcagacggctcctacgctaccaactttgaga 240
44 A I L F D D R Q A P A D G S Y A T N F E 63

241 ccgaggatggcggtgaggggtgtcagagaacggccagcctggatcagctggccagagcaacg 300
64 T E D G V R V S E N G Q P G S A G Q S N 83

301 tagagggatcttatttccttcactgaccctgacggcaacctgggtggaggtgcgctacgtgg 360
84 V E G S Y S F T D P D G N L V E V R Y V 103

361 ctgacgagtttcggcttccgcgctgagagcccctatgtgccacgccccacccctcccg 420
104 A D E F G F R A E S P Y V P T P H P L P 123

421 cccacgccctgcagcagattgcctacgcggaggagctgaggagacttcgtgagcagaacg 480
124 A H A L Q Q I A Y A E E L R R L R E Q N 143

481 gcgagccacagttctcttaatttctaggtgttctctcatacctgctgtacctgttggtgt 540
144 G E P Q F S .

541 tctcctctagtagcttccctacaatacctacatctgttctcaagatgttataatgagctc 600

601 cgaacagcctcgcaaggacccaaaatgtctgctgctgtttgttctttgtgatccttggtgt 660

661 ttgtaataaaatgcctcacgttc

```

## CsAMP16.3

```

1 gcaccgccacacaacatgaagctcggttctcctcgctgcttctcgccctctccgctgcg 60
-15      M K L V L L A C L F A L S A A -1

61 gcccctcagttcctgaggaccagccgtccccagcccttcattcctttcattccccgcgag 120
1 A P Q F L R T S R P Q P F I P F I P P Q 20

121 caaaacaaccgcatcctccccccacagcaaaacaaccgcatcatccaggtgctgagggac 180
21 Q N N R I L P P Q Q N N R I I Q V L R D 40

181 agccgcgatgaccgcggagacggcaactacaactacgagttcgagactgagaatggcatc 240
41 S R D D R G D G N Y N Y E F E T E N G I 60

241 tacaccaacgtggagggagcgactgggctgagcggccagaccaaccaagcaggctccttc 300
61 Y T N V E G R T G L S G Q T N Q A G S F 80

301 aggttcacccttctctgatggcaccctcgctgaggtcaccttcgtagctgatgaggctgga 360
81 R F T L P D G T L A E V T F V A D E A G 100

361 ttccgcgcccagttctccactcctgccccaggccccgcccattgcccgcacgcccctggag 420
101 F R A Q S P L L P Q A P P M P A H A L E 120

421 cacatccgcaaggctgaggaggagcggtgcccggtggcggttcagttcgacagccgaggcttc 480
121 H I R K A E E E R A R G V Q F D S R G F 140

481 agaataggcggttcttaagcttctaggattgtttctccatcacaccagtcgacgcctcat 540
141 R I G G F . 145

541 gactccctcaccctcaccgctgctcgcccccatcaacaccacctgccggcacatg 600

601 aacctacacataaactaccgcccctaaccaccatacacggcggtccatagtaacacacac 660

661 tcacatatatacattgccttgaaggttggtataaaagagct

```

## CsAMP13.4

```

1 ctctccacaccgacaacatgaagctcgtcacctcgctgctggccgccgtcgccctt 60
-15      M K L V I L A C L A A V A L -2

61 gcccgcggccagtcagatccccgaccacatcgaaatcctcagacaggagactgttgacaat 120
-1  A R P Q S D P D H I E I L R Q E T V D N 19

121 ggtgacggaaacttcaactacttgttcgagaccgagaatggcatctacaaggaagtgggt 180
20  G D G N F N Y L F E T E N G I Y K E V V 39

181 ggcagcccatcagccaacggcgcccgagccatgaggggacccctacaggttccccctcgat 240
40  G S P S A N G A Q A M R G S Y R F P L D 59

241 gacggcaccatcggttgaagtcaccttcaccgctgacgagaacggttacctgccccagttct 300
60  D G T I V E V T F T A D E N G Y L P Q S 79

301 gacgccatccccacccccaccctctgcccgtcacgtgatcgagaccctcgccctcgtc 360
80  D A I P T P H P L P A H V I E T L A L V 99

361 gacgagctgggtgagacagggcgccacttgggacgatcaaggaaggagaatcacccgtagg 420
100 D E L V R Q G A T W D D Q G R R I T R R 119

421 aagtaaatatctcacatactaaactcgcgggatgatcatccgcttcgacccttccccgct 480
120 K .

481 gagcacctgccctgctagggctctgtgagcaggagaagtgaaagttactgtcgagtgctaa 540

541 gtggccctacactgtacatggaaacttcgggactgctcgacacctgtaaggaacaaacga 600

601 catccacgtcataaggcgacaccactcactcctgatgtgttgttctttttcccatcacta 660

661 ccaccaccacaccagccaacaccaccacctcacaggtgcttaccctatttacttctcaca 720

721 catacacactttacaatcaagagctttactttcttttttctaacttcacttccgaccacg 780

781 tggccacccatgcagggcacgtcacagcccacacggactcctctcgaccttccaacgtgc 840

841 cacgcttgtagacatcgtaagatctgaagcttcggggctccaaccagcatccctgagcact 900

901 ggacacaatccaatgcttgctcatctcttttttaataaaatagccgcaatc

```

## CsAMP9.3

```

1  ctcatcctcagtaaccaacatgaagctcgccatcttcgctgccttctggcgtcgccct  60
-15      M  K  L  A  I  F  A  C  L  L  A  V  A  L
61  cgcccgccccagggccccaggggatgccagactctccgcagcgagaacatcaacgaggg  120
-1    A  R  P  Q  A  P  G  D  A  Q  T  L  R  S  E  N  I  N  E  G  19
121  caacggcgtcttccagtagaactttgagacagacaacggcattgtagtggaggccagtgg  180
20    N  G  V  F  Q  Y  N  F  E  T  D  N  G  I  V  V  E  A  S  G  39
181  ctcccccaacgctgagggcagcagcgacatcacaggcagctacaaattccctctcgga  240
40    S  P  N  A  E  G  S  S  D  I  T  G  S  Y  K  F  P  L  G  N  59
241  cggacagttccttgaggtgacctacgtttccgacgccagaggattcaggccctccaccag  300
60    G  Q  F  L  E  V  T  Y  V  S  D  A  R  G  F  R  P  S  T  R  79
301  atacgtcagccagaggaagtaagaagcgaggcgacccacacaccgcaactcacacgacat  360
80    Y  V  S  Q  R  K  .
361  gtctccgcatctgaccgggaccacactcacctgaccacaccatgcgctacacctgggata  420
421  tgatgacgcctctcctacacatacaaacacacgatataggaggatggctgtccctacact  480
481  ccgccttccctttcccgctcatttcctttcactttcctcctgattgactccacctccac  540
541  catcttcgcatctcccctgttttctctctgtttcacactgcacactcgtctcgctgtggg  600
601  ttaaaacgagcaccttatccatcgggggtttgttccacgcgatgtccacaatctactgtt  660
661  gttatctcttttgttttttcagcgaataaatgctcttcaat

```

## CsCP6.1

```
1 ctgacacacatacacacaccaacatgaaggtcctgagcgtcatcctgctggccgtggtgc 60
-20 M K V L S V I L L A V V -9

61 tgttcctaggtatgggtcgccgcccgtcccaatgaggtcctggacttcgaaaatgatatga 120
-8 L F L G M V A A R P N E V L D F E N D M 12

121 cgagccacgagcagtagcggtgccccggcactgccgtgcacggcgagtagcaggcaaagg 180
13 T S H E Q Y G V P G T A V H G E Y E A K 32

181 acgccttcggcaactggtacaaggtcaagtacatcgctgacgcaggtggtttcgtgtcg 240
33 D A F G N W Y K V K Y I A D A G G F R V 52

241 tatcgtagacagtgaggacaccaacgcctgctgaggcagcccaacgcgactcgtaaagca 300
53 V S .

301 ggccctccgccaacaacaccggcgacaccgacctcatgtttacctttcctgctcatctctt 360

361 actccttctttctatatatgtaattattcttgaaaataaattctacgtacattt 416
```

## CsCP19.0 top A, bottom B

```

1  ttcaactccttcgtgcccaactcacctactactcctccttctccacttacggcaacatga 60
                                                                M -15
61  gggctctgggtgttcttgcagttgtggcgcatgctccgcatgcctttcatcccgacg 120
-14 R A L V V L A V V G A C S A M P F I P D 6
121  caccagatgtggcggtgagaaaagctcgtttcttccaagcttaccagggttgctcacgccg 180
7 A P D V A A E K A R F F Q A Y Q V A H A 26
181  ccaatctgccaaaggcttcccgaccaagccatcaagcctttgtgccccaccaggccttca 240
27 A N L P K A S R P S H Q A F V P H Q A F 46
241  accaggccccagttcagaccctaagtggatggggcctcttgcttccaacgtccctgctg 300
47 N Q A P V Q T P K W M G P L A S N V P A 66
301  gccttccccggttcctccgctttcgtggccgataactcctgaggtgcaggctgccaagaatc 360
67 G L P G S S A F V A D T P E V Q A A K N 86
361  atttctgaacgcctacagcgcccaggttgccgccaccgtccccgctcggtcccaggactt 420
87 H F L N A Y S A Q V A A T V P V G P R T 106
421  ctacgccatctttcgctccccgcccagttcttccagggttccgcccctgctcctcgcccc 480
107 S Q P S F A P R P V F Q A S A P A P R P 126
481  agcccaagtggaccggtcccccttgcttctcagggtccccgctgggtctgcctggatcttccc 540
127 Q P K W T G P L A S Q V P A G L P G S S 146

a
541  ctgttctctccgacaccgctgaggttgctgctgccaagaacgctttcttccacacctaca 600
147 P I L S D T A E V A A A K N A F F H T Y 166
V
601  gcgctcaggtcgagccaccgcgggagcgccatccaccagggttcttctaagccgcactgt 660
167 S A Q V A A T A G A P S T R F F . 182
661  gataaccctttcaaatttagtgtaattcttctatttcttctatgtcaataaaaccactaat 720

721 agt

```

## CsCP15.0

```

1  gcttttagctaccactccagcaacacactaacatcatgagggctctggttgtcttggcagt  60
-15                               M R A L V V L A V -7

61  gctgggcgcttgctcggcctttaccagtcacccctgatgaccctctagtggcgggcgagaa 120
-6  L G A C S A L P V I P D D P L V A A E K 14

121 ggctaagttcttcgcagcctaccaggctgcactgccagccactccaccaagccagctga 180
15  A K F F A A Y Q A A L P A T P P K P A D 34

181 tccccccaagtgggtacgggcctctggttcttcgcgtccccgctggcctccccggttcttc 240
35  P P K W Y G P L A S S V P A G L P G S S 54

241 ccccggtggtggccccccactgctgaagtggcagctgctcgtaacgagttctacagcaccta 300
55  P V V A P T A E V A A A R N E F Y S T Y 74

301 caacgcccaggtagctgctggttgcgccctaccgcccgtggcccagttcttcaaggtggtgcc 360
75  N A Q V A A V A P T A G G P V F K V V P 94

361 agctgtagttaccgggacatggaccggtcccggtggcagccactattcctgctggtctccc 420
95  A V V T G T W T G P V A A T I P A G L P 114

421 aggctctgcccctaacgtggccgacactgccgacgttgcgaggagccaagaccgctttctt 480
115 G S A P N V A D T A D V A G A K T A F F 134

481 cgacacctaccagagggcaggttgctgccgtggcaccggcccccaaggtgtaagccgcact 540
135 D T Y Q R Q V A A V A P A P K V . 150

541 gtgatacgcttaagattaattcggccaattggatacttgtaacattattgaggaaattcc 600

601 caataaacgaaaacgaatttc

```



## CsCP14.1

```

1 ccacagcactcagcatcatgcacttttgtagattatcgccctcctggcctccgtggccctcg 60
-15 M H F V I I A L L A S V A L -2

61 ccgccccacatatgactcctctagcgagaaggaagcgcccatcctgaagtacgaccgca 120
-1 A A P T Y D S S S E K E A P I L K Y D R 19

121 tccaagaggacgacggcaggtacaactatgagttcgagactgggaacggcatcagccact 180
20 I Q E D D G R Y N Y E F E T G N G I S H 39

181 cccagtcgggtccacggggagacgagaccggcgccatcaacaaagccggacattacacgt 240
40 S Q S G S P G D E T G A I N K A G H Y T 59

241 acacggctcctgacggcaccgtggtagaaatgaagttcgtggccaacgagaatggtttcc 300
60 Y T A P D G T V V E M K F V A N E N G F 79

301 aacccgagtcctcctctgctccctgtggcccccgagttccccaccctatccccagttcg 360
80 Q P E S P L L P V A P E F P H P I P Q F 99

361 tgctagaccagatcgccctttgcagctgagcaggacgcccgcgcccgcggtgagctca 420
100 V L D Q I A F A A E Q D A A R A R G E L 119

421 gcagcgaggagggctcctacagttacaactaatcttcaagccgctaattaggtgactcct 480
120 S S E E G S Y S Y N . 129

481 gccggcgctgataatcttgagtagtgtagtcatatattttatacgaatgagtgagtga 540

541 tggcctgagtgctgggcaaatataagcatgaatctc

```